

Problem Set 2
HEB 1333
Fall 2012

Note: Much of what you need to know is covered in Nunn & Altizer Ch. 4, and will be presented in lecture on 10/02. You should feel free to run the R code in advance of reading or lecture, and deeper understanding will come from the reading and lecture.

HEBV has run its course in our experimental population (HEB 1333 class members and staff) and now it is time for us to analyze the data that we have been able to collect. Your conclusions about the epidemiology of this virus could be crucial to understanding and further controlling this virus. All of the data relevant to the spread of this disease are included in two data files to be downloaded with this problem set ("epidemiology_data.csv" and "spread_data.csv"). All analyses should be conducted in R, as it has become an industry standard of sorts for the types of analyses that we'll be doing. Also download all of the R script files to go along with this problem set ("epigraph.r", "var1_est.r", "var2_est.r", and "spreadvars_graph.r"). Remember before starting any of your work for this problem set to change your working directory (hint: `setwd("<folderpath>")`) to the folder on your computer where you have placed the data files and R script files.

1.a) We found over the course of the experimental introduction that HEBV appeared to have no incubation period (i.e. infected individuals were immediately infectious). Also, we found that after it had run its course of infection, HEBV conferred lifelong immunity. This sort of a parasite can be conceptualized through what type of basic epidemiological model?

1.b) Briefly describe how this type of model operates in 3 sentences or less:

2.a) We have all of the necessary data in our first file, "epidemiology_data.csv" to graph just this sort of an epidemiological model (the one that you described in question 1). Import the data into R (hint: `read.csv()`), and name the imported dataset `epidat`. Then run the R script file by typing `source("epigraph.r")`. Print out the resulting graph (or attach it to an email if submitting electronically).

2.b) What do each of the lines on the graph represent?

Black dashed line: _____

Red solid line: _____

Blue dotted line: _____

Name _____

2.c) Briefly describe what is occurring in this graph:

3.a) We can see from the graph in question 2 that the disease spread rapidly in the beginning of the study. Spread of a disease is often conceptualized as the number of new infections that result from one infectious individual in a population composed of entirely susceptible individuals. What is this number more commonly known as?

3.b) For what values of this number would you expect a parasite to spread through a population?

And for what values would you expect a parasite to NOT be able to spread through a population?

3.c) Judging from the results observed in your work from questions 1 and 2, in what general range do you would you predict that this number would fall for HEBV?

3.d) Calculate this number for the observed spread of HEBV in our class. We have all of the necessary data for this calculation in our other file, "contact_data.csv". Import the data into R, and name the imported dataset `cont`. Then run the R script file by typing `source("var1_est.r")`. What value did you get for this number (hint: type `var1_est` into the command line after running the script)?

How does this value compare to your prediction from question 3.c?

4.a) Luckily we decided to have a controlled release of HEBV into a finite and diminishing population of susceptible individuals, specifically, this class. In such a case, the number described in question 3.a is not really very practical; instead, what is the name of the variable that is more commonly used to describe the number of new infections that result from one infectious individual in a finite and diminishing population of susceptible individuals?

Name _____

4.b) Calculate this number for the observed spread of HEBV in our class. Since we've already imported the data for this estimation, just run the R script file by typing `source("var2_est.r")`. How many value(s) did you get for this number (hint: type `var2_est` into the command line after running the script)? What are the first and last values? Why are there so many estimates for this number?

4.c) Since there are so many values for this second variable, it might be useful to visualize these numbers in a graphical format. Graph this variable compared to the variable in question 3 over time by running the R script `source("spreadvars_graph.r")`. Print out the resulting graph (or attach it to an email if submitting electronically). --- ***On your print-out or electronic copy, give the y-axis an appropriate title based on what the graph is displaying.***

4.d) What do each of the lines on the graph represent?

Black dashed line: _____

Purple solid line: _____

4.e) Why do the two lines start at the same value? Why do we see the purple-lined value dip below the black-lined value, and eventually reach 0?

5.a) A vaccine to HEBV has been developed by the HEB1333 Labs Inc., but it is quite expensive. We'd like to keep any future outbreaks under control, with the Boston metropolitan area (population: 4.5 million) being our primary concern, but we don't want to produce more vaccines than we need to. In order to control an outbreak, we need to know how many individuals in the Boston area to immunize against HEBV. *How many vaccinations* would you suggest for HEB1333 Labs Inc. to prepare for this outbreak; show your work, using the formula presented in class (or the book).

5.b) Explain your reasoning as to why you think this vaccination strategy would work, including a discussion of basic reproductive number and how it relates to determining the proportion of a population to immunize. Is this a minimum or maximum % of the population that should be vaccinated?

Name _____