Mathematics 4MB3/6MB3 Mathematical Biology

http://www.math.mcmaster.ca/earn/4MB3

2019 ASSIGNMENT 2

Group Name: The Plague Doctors

Group Members: Sid Reed, Daniel Segura, Jessa Mallare, Aref Jadda

This assignment was due in class on **Monday 4 February 2019 at 9:30am**.

1 Plot P&I mortality in Philadelphia in 1918

(a) Confirm that you have received this data file by e-mail:

pim_us_phila_city_1918_dy.csv

This plain text comma-separated-value file can be examined (if you wish) using any plain text editor, such as Emacs.

(b) Read the data into a data frame in R, using the read.csv() function. For example, the following chunk of R code should work:

```
datafile <- "../questions_files/pim_us_phila_city_1918_dy.csv"
philadata <- read.csv(datafile)
philadata$date <- as.Date(philadata$date)
opts_chunk$set(dev = 'tikz')
```

The purpose of the last line of code above is to ensure that R encodes character strings such as "1918-10-15" as dates.

(c) Reproduce the Philadelphia 1918 P&I plot:

q1c

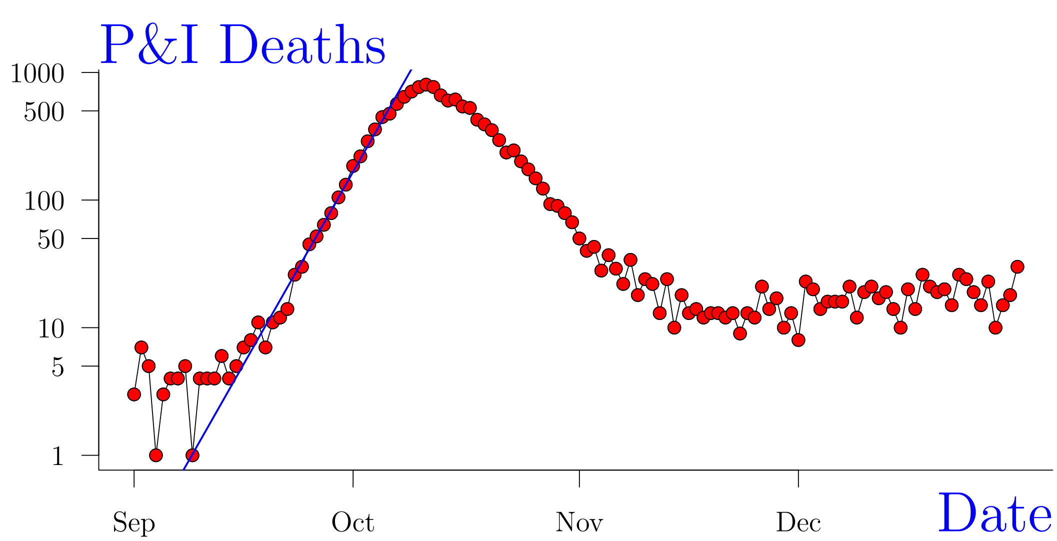
2 Estimate $\mathcal{R}_0$ from the Philadelphia P&I time series

(a) The observed mortality time series $M(t)$ is certainly not equal to the prevalence $I(t)$ that appears in the SIR model. Suppose, however, that $I(t) = \eta M(t - \tau)$ for all time (where η and τ are constants), *i.e.,* that the mortality curve is exactly a scaled and translated version of the prevalence curve. Prove that if both I and M are growing exactly exponentially over some time period then their exponential rates are identical. Thus, if we compare them during the "exponential phase" on a logarithmic scale, then both curves will be perfectly straight with exactly the same slope.

q2a

(b) Fit a straight line to the part of the Philadelphia 1918 mortality time series that looks straight on a logarithmic scale (and show your result in a plot). Once you get the hang of it, the easiest way to do this is to use the `lm()` function in R (lm stands for linear model). Note that the simplest way to draw a straight line with given slope and intercept is with the `abline()` function. If you find `lm()` counter-intutive to understand then experiment with `abline()` until your eyes tell you that you have discovered a line that provides a good fit.

```
date <- philadata$date
pim <- philadata$pim
logpim <- log10(philadata$pim)
#linear model fitted on data points 15 to 35 (looks linear)
line <- lm(logpim[15:35] ~ date[15:35], data=philadata)
#plotting setup
par(las=1, bty="l", mai=c(1,.7,1,.7))
plot(date, pim, log="y", ann=FALSE, xaxt = "n")
#plotting lines and red points
lines(date, pim); points(date, pim, pch=21, bg="red")
#adjusting the axis and labels
xaxis <- seq(as.Date("1918-09-01"), by="months", length.out = 4)
axis(1, xaxis, format(xaxis, "%b"))
mtext("P\\&I Deaths",side=3, line=0, adj=0,col="blue",cex=2)
mtext("Date",side=1,line=1,adj=1,col="blue",cex=2)
#overlaying line of best fit
abline(line, col="blue", lwd=2)
```



(c) How is the slope of your fitted line related to the parameters of the SIR model? (*Hint:* When I is small, $S \simeq 1$.) Why do you need an independent measure of the mean infectious period to estimate $\mathcal{R}_0$? If the mean infectious period is 4 days, what is your estimate of $\mathcal{R}_0$?

```
#Slope of the linear regression line
line <- lm(log(pim[15:35]) ~ date[15:35], data=philadata)
line


##
## Call:
## lm(formula = log(pim[15:35]) ~ date[15:35], data = philadata)
##
## Coefficients:
## (Intercept)  date[15:35]
##   4339.8659       0.2316
```

The slope of the fitted line is 0.2316. In the SIR model, this value corresponds to the initial growth rate, $\beta - \gamma$ because if $S \sim 1$ initially, then

$$\begin{aligned}\frac{dI}{dt} &= \beta SI - \gamma I \\ &\approx (\beta - \gamma)I\end{aligned}$$

An independent measure of the mean infectious period, $(\frac{1}{\gamma})$, is needed to estimate $\mathcal{R}_0$ because $\mathcal{R}_0$ is the product between $\frac{1}{\gamma}$ and β, the transmission rate.

To calculate an estimate of $\mathcal{R}_0$ given $\frac{1}{\gamma} = 4$:

$$\begin{aligned}\gamma &= 0.25 \\ 0.2316 &= \beta - \gamma \\ \beta &= 0.2316 + 0.25 = 0.4816 \\ \mathcal{R}_0 &= \frac{\beta}{\gamma} = (0.4816 * 4) = 1.9264\end{aligned}$$

The estimated $\mathcal{R}_0$ is 1.9264.

3 Fit the basic SIR model to the Philadelphia P&I time series

(a) Install the `"deSolve"` package. This is done by typing the following command in the Console pane of `RStudio`:

```
install.packages("deSolve")
```

You will then be prompted to choose a mirror site from which to download the package. It doesn't matter which mirror you choose, but choosing a site in Ontario might save a fraction of a second. *Note:* This is a one-time operation. You do not want an `install.packages()` command inside your solutions code.

(b) Write an R function that plots the solution $I(t)$ of the SIR model for given parameter values ($\mathcal{R}_0$ and $1/\gamma$) and given initial conditions (S_0, I_0). Use the `ode()` function in the `deSolve` package.

q3b

(c) For $I_0 = 10^{-3}$ and $S_0 = 1 - I_0$, plot the solutions of the SIR model assuming $1/\gamma = 4$ days and $\mathcal{R}_0 \in \{1.2, 1.5, 1.8, 2, 3, 4\}$. Use the `legend()` command to make a legend on the plot that shows which curves correspond to which values of $\mathcal{R}_0$.

q3c

(d) By trial and error, find values of $\mathcal{R}_0$ and γ that yield a solution of the SIR model that fits the Philadelphia P&I times series reasonably well. You can assess the quality of fit using the Euclidean distance between the model solution and the data. (*Note:* The trial and error approach is a valuable exercise, but not a suggestion of a method you would really use in practice. We'll discuss better methods for fitting ODE models to data later.)

q3d

4 Executive summary for the Public Health Agency

q4

— END OF ASSIGNMENT —

Compile time for this document: February 2, 2019 @ 22:30