Purpose: a) Learn about disease progress curves; b) Distinguish among some common disease progress models; c) Calculate rates of disease increase, and use these for various manipulations; d) Compare two epidemics; e) Calculate AUDPC.

Below are two disease progress curves. These are observations of incidence of *Cucumber mosaic virus* (CMV) and *Watermelon mosaic virus-2* (WMV-2) in a melon crop in Spain. Details are in Alonso-Prados et al. (*Euro. J. Plant Pathol.* 109: 129-138 [2003]). Both viruses are of the polycyclic type, with spread from plant to plant (through their aphid vectors). The paper shows results for several years and locations, but the data here are for one year and location (see Fig. 1J in article). I actually read the values from the graph, so there is likely some inaccuracy, especially at low y values. Just assume that the listed points are correct.

Note: the authors conclude that the Gompertz model is best (of those evaluated) overall for the many epidemics. This does not mean that the Gompertz is best for the two specific epidemics here.

<table>
<thead>
<tr>
<th>t</th>
<th>y (CMS)</th>
<th>y (WMV-2)</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0.01</td>
<td>--</td>
</tr>
<tr>
<td>7</td>
<td>0.05</td>
<td>--</td>
</tr>
<tr>
<td>12</td>
<td>0.17</td>
<td>0.01</td>
</tr>
<tr>
<td>19</td>
<td>0.36</td>
<td>0.03</td>
</tr>
<tr>
<td>26</td>
<td>0.71</td>
<td>0.09</td>
</tr>
<tr>
<td>33</td>
<td>0.86</td>
<td>0.28</td>
</tr>
<tr>
<td>40</td>
<td>0.93</td>
<td>0.33</td>
</tr>
<tr>
<td>46</td>
<td>0.95</td>
<td>0.59</td>
</tr>
</tbody>
</table>

Another note: the authors use \( t = 0 \) for date of transplanting. However, I use \( t = 0 \) for 30 days later when disease symptoms are first seen (at least for CMS). Use the listed times here as the actual ones of the epidemics (i.e., do not change times for WMV-2 epidemic). For WMV-2, the first observed or reported time is \( t = 12 \). In other words, there are only six different times for WMV-2 incidence measurements, and the first observation is \( t = 12 \) and \( y = 0.01 \). With this convention, your estimated intercept (below) will be the transformed disease intensity at \( t = 0 \) for both epidemics (30 days after transplanting, which is what you want). You will see that \( y_0 \) is lower for WMV-2.
For each data set:
1) Plot $y$ versus $t$ (show points, and connect the points by lines).
2) Plot estimated $\frac{dy}{dt}$ versus $t$ (you can estimate $\frac{dy}{dt}$ as $\frac{\Delta y}{\Delta t}$ with a calculator).
3) Plot a collection of $y^*$ (e.g., $\ln(y)$, etc.) values versus $t$. Assume maximum $y$ is 1.0.
4) Decide on an appropriate disease progress model (choose between exponential, monomolecular, logistic, and Gompertz). Base this on the above plots and regression analysis (linear regression). Give reasons for decision. Caution: your conclusions can be much stronger for the CMV epidemic compared to the WMV-2 epidemic, because the $y$ values did not approach 1.0 for the latter epidemic. Thus, only some of the graphical methods for selecting a model are useful for WMV-2. It is possible that you will find that the best model is not the same for each epidemic. Given the strength of the evidence, select which SINGLE model (exponential, monomolecular, Gompertz, or logistic) is best (overall) for both of the epidemics. (You do not need to consider the Richards model directly). You should only be using ordinary linear least squares regression (no weights).
5) For your choice of the best model:
   1) Show the parameter estimates ($r^*$, $y_0^*$) from ordinary least squares regression analysis, together with the standard errors. Show appropriate regression output, including residual plots.
   2) Calculate an estimated 95% confidence interval for $r^*$. See lecture notes or equation 3.14 in the textbook (Chapter 3) for the generic approach to obtaining the confidence interval for a model parameter. This requires use of critical the Student $t$ value (i.e., the upper 97.5-th percentile of the Student $t$ distribution, when 95% confidence is desired). For the CMV epidemic, there are 8 data points. This means that df=6. Use a $t_{0.05/2,6}$ value of 2.45. For the WMV-2 epidemic, there are 6 data points. This means that df=4. Use a $t_{0.05/2,4}$ value of 2.78. It is possible that your chosen statistics program will automatically do these calculations (for confidence intervals)—you still must show how to do this “manually”.
   3) Now calculate a 95% confidence interval for $y_0$. Note: you must actually determine the confidence interval for estimated transformed $y_0^*$ (which is the linear regression-model parameter for the intercept). Then you back-transform the two endpoints of the $y_0^*$ confidence interval to get the $y_0$ confidence limits. The specific backtransformation depends on the model chosen (be careful). You can use the table of transformations for this. The same critical Student $t$ values are used for the intervals as with #2 (above).
   4) Based on the chosen model, what is the implied shape parameter ($\eta$) of the Richards model? What is the calculated $\varpi$ value (see lecture notes or equation 4.26 in Chapter 4) based on the implied shape of the Richards model?
   5) Are the two epidemics different in terms of their estimated rate parameters?
Determine this with a $t$-test as described in class or on page 94 of the textbook (Chapter 4). Degrees of freedom would be $8 + 6 - 4 = 10$, giving a critical $t$ value of $t_{0.05/2,10} = 2.2$ (for 95% confidence)—in other words, if the achieved (calculated) $t$ statistic exceeds 2.2 in absolute value, then reject the null hypothesis.

6) **BONUS (optional):** how does your estimate of $r^*$ change if you used weighted linear least squares, where the weights are given in Chapter 4 based on the theoretical variance of a proportion?

NOTE: USE THE OVERALL BEST MODEL FOR THE ABOVE MULTI-PART ANALYSES IN #5.

6) Using the estimated parameters from regression analysis (and the assumed best model overall [for both epidemics]), calculate the doubling time from $y = 0.40$ to $y = 0.80$ (do this last part with a calculator) for the WMV-2 epidemic. Explain the relationship between doubling time and $r^*$. Explain why doubling time depends on $y$ (when $y > 0.10$ or so). (Be careful, the doubling-time calculation depends explicitly on the chosen model, and is only independent of the starting $y$ value for the exponential).

7) Using the estimated $r^*$ value (e.g., $r_L$, $r_G$, etc.) from regression analysis for the CMV epidemic (and the best model overall), determine how much time would be gained when initial disease intensity is reduced 10-fold (from 0.01 to 0.001) through some “sanitation” method (do this with a calculator, after getting $r^*$ from computer program for regression analysis). In other words, determine the time it would take for $y$ to increase from 0.001 to 0.01 at the same $r^*$. (Note: estimated $y_0$ from the model fit will not necessarily equal the observed $y_0$; use 0.01 as the initial disease without sanitation in this calculation). Caution: make sure you use the sanitation-time formula specific to your chosen disease progress model (don’t assume exponential).

8) Calculate AUDPC for the CMS epidemic using eq. 4.42 in book (or from notes in class), from $t=0$ to $t=46$ days. Do this with the observed $y$ and $t$ values, not with predictions of $y$. **BONUS (optional):** calculate AUDPC for the WMV-2 epidemic from $t=0$ to $t=46$. For times 0 and 7, use estimated $y$ based on the regression analysis (with your selected best model), and observed $y$ values for the last six times. That is, using your estimates of $r^*$ and $y_0^*$ for this epidemic, predict the values of $y^*(0)$ and $y^*(7)$, and use backtransformation to get $y(0)$ and $y(7)$ [of course, the intercept from the regression analysis automatically gives you $y^*(0)$]. Then you have a “complete” disease progress curve with the same number of time points as with CMV.

*There are many calculations here, so it is important to show enough of your work so that we can tell if you are doing the calculations properly. A small numeric mistake at the beginning could lead to incorrect...*
final results. If you do not show some of the intermediate steps, we cannot tell if you understand the procedures. Results should be organized in tables, and narrative answers should be in paragraph form. Graphic output from a computer program can be pasted into the document. Your assignment report should be submitted as a document in Microsoft WORD. This assignment is worth more than double the first assignment.

Send WORD file to Lee Wilson (Wilson.40@osu.edu).