BIOL811 Applied Biostatistics II

Non-linear Regression in JMP

Recap: For studies with a quantitative experimental factor, you learned how to fit a straight line that minimizes the sum of the squared deviations (SS\text{Error}) between the line and the data points. You calculated the SS\text{Treatment} by adding up the squared distances between the overall mean and points on the line corresponding to each data point. From the SS\text{Treatment} and SS\text{Error}, you were able to calculate the MS\text{Treatment} and MS\text{Error} and an F-ratio (MS\text{Treatment}/MS\text{Error}). The p-value associated with that F-ratio is the probability the treatment has no effect on the response. You also saw that it is possible to fit a straight line through data and get a p-value less than 0.05 even when the relationship between the treatment and response is clearly not linear. A Simple Linear Regression is not a test for linearity.

Objective: In studies where the relationship between the experimental factor and the response can be represented by a curve, and where a mathematical formula for the curve can be defined, it is possible to fit that curve to the data. This is done by adjusting the parameters of the formula to minimize the sum of the squared deviations between the curve and the data points (SS\text{Error}). This is analogous to adjusting the slope and intercept of a straight line to find the best fit.

The first objective of this exercise is to use JMP to fit curves to data. The second is to determine whether the curves fit through two sets of data are similar or significantly different.

Activity 1: On the Bayberry PS Data sheet of the Non-linear.xlsx workbook are the results of a field/lab study on the effects of light level on the photosynthetic rate of bayberry plants (\textit{Myrica pensylvanica}) that are acclimated to full sunlight or to shade. Up to a point, the photosynthetic rates of plants increase with greater light. A mathematical model that often can be made to fit photosynthesis vs light data fairly well is as follows:

$$P_S = \frac{P_{\text{max}} \cdot \text{Alpha} \cdot \text{Light}}{P_{\text{max}} + \text{Alpha} \cdot \text{Light}}$$

Where $P_S$ is the photosynthetic rate (mgO$_2$/m$^2$/h), Light is light level ($\mu$moles/m$^2$/s), $P_{\text{max}}$ is the maximum photosynthetic rate (when the plant has all the light it can use) and Alpha is the initial slope of the curve. Alpha is an indication of how the plant responds in low light levels.

Open JMP and close the Tip-of-the-Day window. Pull down File -> New -> Data Table. Copy and Paste the 3 columns of data (not the headings) from the Bayberry PS Data worksheet of ANCOVA.xlsx. You should have three columns of data. Double click on where the headings and name the columns Light, P$_S$ _Sun and P$_S$ _Shade. Right-click where the heading would go for a fourth column and select “New Column...”. Change the column name to P$_S$ vs Light and click OK. Right-click on the PS vs Light column heading and select Formula. This will open a dialog box. Pull down the Constants menu in the lower left of the dialog box and select Parameters. Click on New Parameter... Enter “P$_{\text{max}}$” as the name and 300 as its value. Click OK. Click on New Parameter again and enter “alpha” as the name and 1 as its value. Click OK. Double-Click on the box that says
No Formula. Click on Pmax in the Parameters box. Press the * key. Click on alpha in the
Parameters box. Press the * key. In the Columns box click on Light. In the formula that
is being created, type a right parenthesis. Press the / key. This should put everything
you created so far on the top half of a fraction. Press the ( key. In the Parameters box,
click on Pmax. Press the + key. Click on alpha in the Parameters box. Press the * key. In
the Columns box, click on Light. Press the right parenthesis key. Click OK. The PS vs
Light column (Column 4) should fill with values that are more or less like the values in
the PS_Sun and PS_Shade columns.

Pull down the Analyze menu and select Specialized Modeling -> Nonlinear. Put PS_Sun in
the Y, Response box and put PS vs Light in the X, Predictor Formula box. Click OK. It
will draw a curve through the data using the parameter estimates that you provided
(Pmax=300 and alpha=1) when you made the formula for the PS vs Light column. To have
JMP find the parameter values for the best fit curve, press the Go button. JMP will
iteratively adjust the Pmax and alpha values to minimized the SSError and then it will
replot the data and curve using its final parameter estimates. Record the Pmax and
Alpha estimates and the SSError value in the spaces provided on the "Curve Comparison"
worksheet of the Non-linear.xlsx workbook.

Print your JMP output to hand in. Be certain it shows all the results that were needed to
complete the exercise (e.g. the PMax, Alpha, SSError, N)

Repeat the analysis for the photosynthesis measurements from the shade acclimated
plants. Copy the parameter estimates for Pmax and Alpha and the Residual SS (SSError)
and paste them in the spaces provided on the "Curve Comparison" worksheet of the Non-
linear.xlsx workbook.

Print your JMP output to hand in.

Activity 2: From the parameter estimates for the Sun and Shade acclimated plant data, it should
be apparent that the curves are different. In other word, the light under which the
plants are grown affects it photosynthesis vs light relationship. In other instances, this
difference may not be so obvious. In those cases, we can do an F-test to test the null
hypothesis that the two curves are not really different. To see how this works, we will
do the analysis on the two curves we just fit. The logic of the test goes like this: if the
photosynthesis versus light response of the shade and sun adapted plants is the same,
then the amount of variability explained by fitting two separate curves to the data will
be no greater than the amount explained by fitting one curve through all of the data.
So what we need to do now is fit a single curve through the sun and shade data
combined.

Click on the Data Table in JMP. Copy all of the Light data and paste it at the bottom of
the Light column (starting at row 25). You should now have two copies of the Light
values. You will notice that the predicted values in rows 25 through 48 of Column 4 (the
Formula column) are automatically filled in. You will also see that the same rows in the
PS_Sun and PS_Shade columns are filled in with dots, which indicate there is no data in
these cells. Copy all of the data from the top half of the PS_SHADE column and paste it
into the bottom half of the PS_Sun column (starting in row 25). You will need to highlight all of the cells in the PS_Sun column that contain dots before pasting. Double-click on the PS_Sun heading and edit it to say PS_ALL. Right-click on the heading of the PS_SHADE column and select Delete Column.

From the Analyze menu, select Modeling->Nonlinear. Put PS_ALL in the “Y, Response” box and PS vs Light in the “X, Predictor” box. Click OK. Press GO on the Nonlinear Fit dialog page. JMP will fit a single curve through all of the data in the combined sun and shade data sets.

Copy the parameter estimates for Pmax and Alpha and the Residual SS (SS_{Error}) and paste them in the spaces provided on the "Curve Comparison" worksheet of the Non-linear.xlsx workbook.

Print your JMP output to hand in.

Activity 3: The "Curve Comparison" worksheet contains a template for an F-Test to determine what is referred to as “the coincidence of two regression lines”. Essentially we are going to test the null hypothesis that fitting two separate lines through the photosynthesis versus light data will not explain any more variability than a single line through both sets of measurements. It also includes a scatter plot to help visualize the analysis.

You have already filled in the Parameters and SS_{Error} table using the data you recorded from the JMP Nonlinear Regressions. The row labeled “n” is the number of data points in the respective data sets. Fill in the values for n. Below the table, fill in the value for m, which is the number of parameters that were fit in each regression. In this case it was Pmax and Alpha, so m=2. Fill in the box for k, which is the number of curves you are comparing, which would be the Sun curve and the Shade curve.

Below this is what at first may appear to be a daunting formula, but if you look at each term separately, it’s really not all that complicated. The top line is \( SSE_{All} - (SSE_{Sun} + SSE_{Shade}) \). This is the amount of variability accounted for by splitting the data into two groups (Sun and Shade) compared to lumping it all together. Do this calculation in cell F28. All you need to do is add the Sun and Shade SSE values from your Parameters and SS_{Error} table and subtract them from the SSE value in the column labeled All, so the Excel equation in F28 will be \( =F22-(C22+D22) \). In F29, you will calculate the df for the numerator of the F-ratio, which is \((m+1)*(k-1)\). If you enter the formula correctly, you should get 3. F30 is the sum of the Sun and Shade SSE values. Finally, F31 is the df for the denominator of the F-ratio. Finish the F-ratio calculation in H29 using \( =\frac{(F28/F29)}{(F30/F31)} \). Use the FDIST function to calculate p in cell F33.

**Question 1:** Are the two curves significantly different?

**Question 2:** Do Sun and Shade adapted Bayberry plants respond differently to increasing light?

Print the Curve Comparison page of your Non-linear.xlsx workbook to hand in.

Assignment: Hand in a printout of your completed MS Excel workbook and your Activity 1 and Activity 2 JMP output.