CALCULATING FITTED VALUES FROM A Y BY X SPLINE FIT

By Annette Sanders
SAS Institute Inc.

To fit a spline relationship to two numeric variables you first use the Fit Y by X command from the Analyze menu. This example uses the BIG CLASS data table from the SAMPLE data, with height as Y and weight as X. The Fit Y by X platform begins by showing a scatterplot of the X, Y data points. The Fitting popup menu beneath the plot accesses the fitting options shown in **Figure A**.

The Fit Spline option fits a smoothing spline using a smoothing parameter you specify. The spline is displayed on the X, Y scatterplot (weight and height in this example) and a table appears showing the R-Square and Sum of Squares Error.

The Save Predicteds popup menu command for the spline fit creates a new data table column and saves predicted values for each row. However, the spline does not have a prediction equation so you cannot find *fitted values* (predicted values) for data points that are not in the data table.

In order to compute fitted values you need more information about the spline fit; specifically, the coefficients of the spline's prediction formula. To get these you first use the Output Coef Table popup menu command on the Spline Fit table, which creates a new JMP table and saves a set of spline coefficients for each unique value of the X variable. **Figure B** shows the coefficients table for the height by weight example.

Note that although the BIGCLASS data table has 40 observations the weight (X) variable has only 29 unique values (29 rows) listed in the O column. These values are called *knot points.* The knot points are points at which third degree polynomials are spliced together. The polynomial values and their first derivatives agree at these points, which results in a continuous and smooth curve.

You can use this new table and the JMP calculator to build a formula that will compute fitted values for any X value. The way you do this is to generate the X values you want, and create an index that associates each X with the appropriate row of spline coefficients for that X value.

The fitted values are for any number you choose of equally spaced X values that fall within the range for which there are spline coefficients (the min and max of the weight variable in this example).

Follow these steps to generate the fitted values.

1) In the spline coefficients table, add the number of rows so the total rows is the number of fitted

Figure A Fitting a Spline Curve

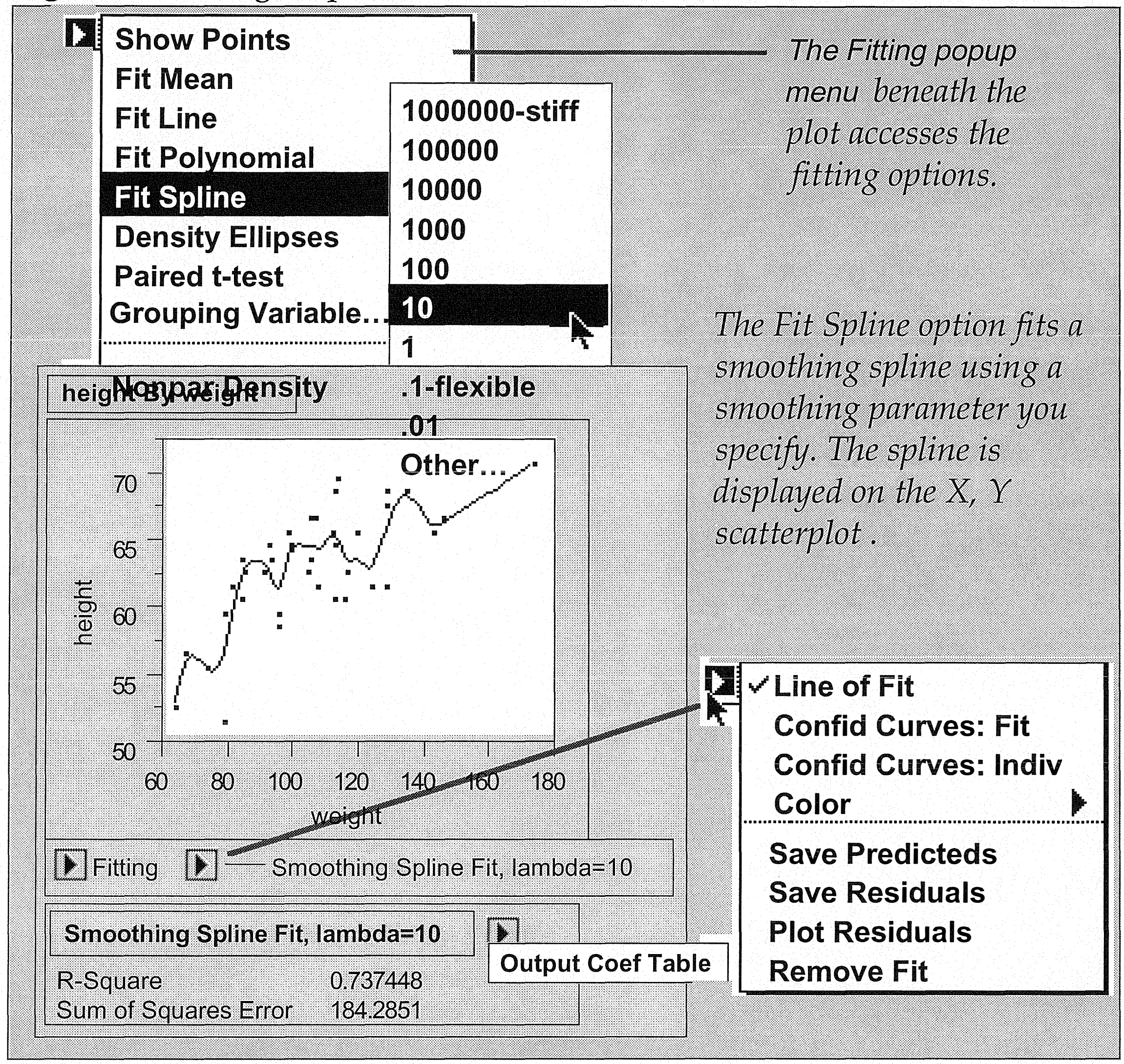

Figure B Table of Spline Coefficients

5 Cols / 29 Rows	O	A	B	C	D
1	64	52.52	1.05	0.00	-0.02
2	67	55.21	0.59	-0.15	0.01
3	74	54.86	-0.25	0.03	0.01
4	79	56.17	1.11	0.24	-0.06
…					
27	142	65.48	-0.08	0.08	-0.01
28	145	65.73	0.16	-0.00	0.00
29	172	70.00	0.00	0.00	0.00
…					

The **O** *column contains unique values of the* **weight** *variable. The columns* **A**, **B**, **C**, *and* **D** *are the spline coefficients associated with each unique* **weight** *value. They are the constant, linear, quadratic, and cubic coefficients respectively.*

points you want to obtain. In this example we added 71 rows (to the 29) in order to obtain 100 fitted points. You can create as many of these new X values as you want at any time by increasing the number of rows in the data table.

2) Use Tables→New Column to create a new column of X values (call it newX). In the New Column dialog select Formula as the Data Source.

3) Use the calculator Count function with Quantile functions as arguments to generate equally spaced values from the minimum to maximum of the knot points:

   count (from $\text{quantile}_0 O$
   to $\text{quantile}_1 O$
   in n steps, 1 time)

   Recall that Quantile_0 of O is its minimum, Quantile_1 is its maximum, and n is the total number of table rows.

   Note that the newX values are in the coefficients data table but their ordering and physical relation to the coefficients are not relevant; you are storing unrelated pieces of information within the same row of the data table. It is likely that the coefficients needed to compute the fitted value for a specific newX value will not come from its own row. The row of coefficients needed is identified by an index variable you need to create, as described next.

4) Create an index column (call it index) with calculated values between 1 and 29 that point to each unique weight value. This index associates each newX value with the appropriate row of coefficients:
   - Use the Sum function to increment the value of index by 1 each time the value of newX is greater than or equal to the next consecutive knot point:

   $$\sum_{j=1}^{n} (O_j \le newX_i \text{ and } O_j \ne \bullet)$$

   - When the value of newX is less than the next consecutive knot point nothing more is added—that index value identifies the row of coefficients needed to calculate the fitted value.

5) Now you can use the JMP calculator to compute fitted height values from newX using the appropriate coefficients. The formula to compute the fitted value from newX with the index value *i* is:

   $$\text{fitted value}_i = A_i + d \bullet B_i + d^2 \bullet C_i + d^3 \bullet D_i$$

   where d is the difference between newX and the observed weight associated with it by its index value. An efficient way to construct this formula is with a temporary variable and an Assignment function. You construct the formula like this:
   - Create a new column called fittedHt.
   - Use the calculator and select New Variable from the Variables

functions list. Create a temporary numeric variable (call it *diff*).

- Select Assignments from the Conditions functions. Use the first clause in the assignment to subtract the knot point (O) subscripted by the computed index (index) from the newX value and assign this difference to *diff*:

$$diff \Leftarrow newX - O_{index}$$
results $\square$

This is done because the coefficients are defined for the increment from the lower end of the known point range containing the newX value.

- The last step is to enter the computation for the fitted value as the clause for results giving the formula and results shown in **Figure C**.

As a visual verification that the fitted values follow the Fit Y By X spline fit you can use Graph→Overlay Plots with fittedHT as Y and newX as X. **Figure D** shows that the computed curve is almost the same as the fitted spline.

Figure C Computations to Produce Fitted Spline Values

$$diff \Leftarrow newX - O_{index}$$
$$\text{results} \quad A_{index} + diff \cdot B_{index} + diff^2 \cdot C_{index} + diff^3 \cdot D_{index}$$

8 Cols / 100 Rows	O	A	B	C	D	newX	index	fittedHT
1	64	52.52	1.05	0.00	-0.02	64.00	1	52.52
2	67	55.21	0.59	-0.15	0.01	65.09	1	53.64
3	74	54.86	-0.25	0.03	0.01	66.18	1	54.64
4	79	56.17	1.11	0.24	-0.06	67.27	2	55.36
5	81	58.82	1.29	-0.15	0.01	68.36	2	55.76

Compute the fitted value for height using the coefficients in the row identified by the index variable.

Figure D Comparison of Spline Fit (left) and Fitted Values (right)