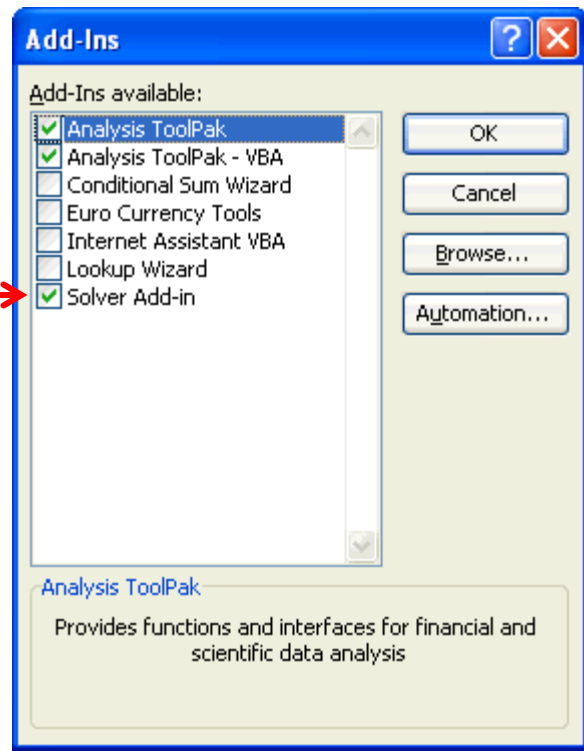


Using Excel's Solver to minimize the sum of squared error for a mathematical model

First the Solver is an add-in that must be loaded. Go to the Tools menu and select Add-Ins... and the screen to the right will pop-up. (For loading Solver in Excel 2007 - [click here.](#))

You must select the Solver Add-in and then press the OK button.



Now you have a tool that allows you to do a number of optimization operations.

The Solver can find the maximum and minimum values, as well as the roots to an equation or a system of equations.

One way to use it is to introduce obtaining the line (or curve) of best fit or regression analysis to students.

In the mathematical modeling of data you are trying to fit an equation to the data. How do you get the best fit of the equation to the data? Regression techniques adjust equation parameters to minimize the sum of the squared errors, SSE. The error is the difference between the measured value of the datum point minus the predicted value of the datum point. The predicted value is calculated from the regression equation. The difference is referred to as the **error** or the **residual**. Here is the column set-up for the data and calculations:

			error	squared error
x	y_{measured}	$y_{\text{predicted}}$	$y_{\text{measured}} - y_{\text{predicted}}$	$(y_{\text{measured}} - y_{\text{predicted}})^2$

The last column is then summed yielding the sum of the squared errors, SSE.

To do this in a spreadsheet, we need to set up the equation parameters to be adjusted manually, yellow cells (no scroll bars or spinners). This gets students playing with the parameters to fit the line/curve manually to try to eye-ball a good

fit. Hence, they need to know or discover how the parameters influence the line/curve. As they adjust parameters, they can observe how the SSE behaves and try to minimize it. Now you can see how well your manual fit is to the best-fit.

Let's consider fitting a quadratic equation ($y = ax^2 + bx + c$) to some data as illustrated on the screen shot below. Go to Tools on the menu bar and select Solver... and the Solver Parameters pop-up menu will appear.

	A	B	C	D	E	F	G	H	I	J	K	L	M
1													
2													
3	x	Y _{measured}	Y _{predicted}	error	squared error								
4	-5	24.00	21.00	3.00	9.00		a =	1					
5	-4	16.00	13.00	3.00	9.00		b =	1					
6	-3	8.50	7.00	1.50	2.25		c =	1					
7	-2	3.90	3.00	0.90	0.81								
8	-1	0.90	1.00	-0.10	0.01								
9	0	0.10	1.00	-0.90	0.81								
10	1	1.10	3.00	-1.90	3.61								
11	2	4.20	7.00	-2.80	7.84								
12	3	8.50	13.00	-4.50	20.25								
13	4	15.30	21.00	-5.70	32.49								
14	5	26.00	31.00	-5.00	25.00								
15													
16					SSE =								
17													
18													
19													
20													
21													
22													

} by changing these cells

SSE = 111.07

This is the target cell and we want to minimize this value by adjusting a, b, and c above.

Click in the **Set Target Cell** box on the Solvers Parameters and then click on cell E16 (the blue cell) on the worksheet, set the **Equal To** to Min (minimum value) then click the **By Changing Cells** box followed by clicking the yellow cells H4, H6, H8 (separated with a comma). Now click on the Solve button. The Solver Results menu will pop-up as shown below. It will default to the "Keep Solver Solution."

Solver found a solution. All constraints and optimality conditions are satisfied.

Reports

- Answer
- Sensitivity
- Limits

Keep Solver Solution
 Restore Original Values

The Solver results will be in the yellow cells for a, b, and c. Click the OK button and you are done. You now have the best-fit regression curve results for your data. Once you have used the Solver to get results on a worksheet and saved the workbook, it will save the appropriate cells in the Solver parameters.

This method allows you to expand non-linear regression techniques that are not available in the trendline types, such as a sine regression on periodic data, logistic regression for population data, or the translated exponential decay ($y = ae^{-kx} + c$) needed for Newton's Law of Cooling data. See the [Non-linear Regressions by Using Solver](#) "Just Add Data" Excelet!

This is based on the article by R. M. Horton and W. H. Leonard (2005) *Mathematical Modeling in Science*, which appeared in **The Science Teacher** 72 (5), 40-46.

If you are comparing different models, the goodness-of-fit can be judged by the model with the minimum SSE as the best-fit. You want to consider a graphical measure of the goodness-of-fit for the regression, plotting the residuals as a function of the x-variable is a good way to do this. This plot should have the residuals distributed in a random fashion. A pattern in the residuals plot is an indication of the mathematical relationship not truly fitting the data, even if the SSE is small.