THE CHINESE UNIVERSITY OF HONG KONG Department of Mathematics

Mathematical Modeling Project Team mathmodel@math.cuhk.edu.hk

Modeling with R Shiny (General Fitting for XY data)

Last updated: 03/07/2025

1 Introduction to the General Fitting (XY) R Shiny tool

R is a powerful statistical programming language that allows us to perform data analysis, which is useful for mathematical modeling. Even better, we have set up a set of online tools using R Shiny so that you can perform various mathematical modelling tasks without any coding background.

In particular, a common task in mathematical modelling is to perform nonlinear regression based on some data points $(x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n)$. Other than our Nonlinear Regression (XY) tool, we have the General Fitting (XY) tool for you to input your own functions for data fitting. Here is the guideline on how to use our General Fitting (XY) R Shiny tool.

2 Using the General Fitting (XY) R Shiny tool

To use the tool, you can go to our main website:

https://www.math.cuhk.edu.hk/app/mathmodel/tool.html

for the list of tools, or simply go to

https://mathmodelcuhk.shinyapps.io/general-fitting/

for the General Fitting (XY) R Shiny tool.

After getting into the website, you can see the following:



Don't worry if you think this is complicated! We will guide you step by step, and you can do nonlinear regression using our R Shiny tool.

2.1 Step 1: Data input

At the top left-hand corner, you can see "Upload your file", this is where we input the data. There are three different buttons.

Firstly, the "?" gives you the general guide on data input.

Secondly, the "+" allows you to input data directly. After clicking it, you can see the following:

Input your data here in csv				
N.V.				
л,у 1 5 40 6				
3.8.65				
89708				
10.78				
12.84				
20 108 5				
20,100.5				

On the first line, you can input the names of the variables for the x-axis and y-axis respectively. For example, we can define the x-axis as "Month" and the y-axis as "Average hours of daylight" to see the correlation between them. Therefore, on the first line we type "Month, Average hours of daylight"

For the following lines, you can simply type in the data points respectively. You should be able to have a result that looks like this:

Input your data here in csv			
	Month, Average hours of daylight		
	1,10.68		
	2,11.3		
	3,11.98		
	4,12.77		
	5,13.42		
	6,14.75		
	7,13.6		
	8,13.05		
	9,12.3		
	10,11.57		
	11,10.88		
	12,10.53		

Scroll down and press "submit". Your data points should be shown on the screen like this:

Gerneral Fitting XY	≡	
General Fitting XY		•
C Linear Regression	Upload your Hiel + 2 Browse No file selected	
Non-Linear Regression	14	
Non Linear Regression (XY)	Input Your Function: ?	
C Multi Regression	×	
Find What Fits	Submit Fit ?	•
Curve Fitting	Best Fit Curve	
	Residuals	· · ·
	Squared Residuals	· · · · ·
	Prediction of:	5.0 /.5 10.0 1
	0	

You can also upload an Excel file to the website directly. Firstly, you will need to prepare your Excel file like this:

	А	В	
1	Time	Hours of Daylight (hrs)	
2	1	10.68	
3	2	11.3	
4	3	11.98	
5	4	12.77	
6	5	13.42	
7	6	14.75	
8	7	13.6	
9	8	13.05	
10	9	12.3	
11	10	11.57	
12	11	10.88	
13	12	10.53	
14			

Then go back to our website, press the "Browse..." button and look for the file stored on the computer. After uploading, you will see your data points shown on the graph.

If you face any problem regarding the upload, you may need to save your Excel file in .csv format. To do so, you can go to file \rightarrow save as, and you can look for the following:



Expand the dropdown list, and look for the .csv format.

test	
Excel Workbook (*.xlsx)	~
Excel Workbook (*.xisx)	Save
Excel Macro-Enabled Workbook (*.xism)	Up save
Excel Binary Workbook (*.xlsb)	
Excel 97-2003 Workbook (*.xis)	
CSV UTF-8 (Comma delimited) (*.csv)	
XML Data (*.xml)	
Single File Web Page (*.mht, *.mhtml)	
Web Page (*.htm, *.html)	
Excel Template (*.xitx)	
Excel Macro-Enabled Template (*.xitm)	
Excel 97-2003 Template (*.xlt)	
Text (Tab delimited) (*.txt)	
Unicode Text (*.txt)	
XML Spreadsheet 2003 (*.xml)	
Microsoft Excel 5.0/95 Workbook (*.xls)	
CSV (Comma delimited) (*.csv)	
Formatted Text (Space delimited) (*.pm)	
DIF (Data Interchange Format) (*.dif)	
SYLK (Symbolic Link) (*.slk)	
Excel Add-in (*.xlam)	
Excel 97-2003 Add-in (*.xla)	
PDF (*.pdf)	
XPS Document (*.xps)	
Strict Open XML Spreadsheet (*.xlsx)	
OpenDocument Spreadsheet (*.ods)	

After that, you can upload the file onto our website, and it will work normally.

2.2 Step 2: Data analysis

After you have inputted the dataset, now we can perform linear or nonlinear regression easily. Try to click different buttons on the page and now we will briefly explain their usage.

Gerneral Fitting XY	=	
General Fitting XY		
C Linear Regression	Browse Hours of Daylight.csv	
Non-Linear Regression	Upload complete	
Non Linear Regression (XY)	Input Your Function: ?	
Multi Regression		*
Pind What Fits	Submit Fit ?	
	Best Fit Curve	· · · · · · · · · · · · · · · · · · ·
	Residuals	n
	Squared Residuals	25 50 75 MD 125
	Prediction of:	×
	0	

If you have any questions about the tool, you can always try to click different boxes with "?" and there will be a short introduction.

2.2.1 Input Your Function

In the box below, you can input the function that you want to use for the data fitting. In our case, since the hours of daylight are periodic throughout seasons, we try to model it with a sinusoidal function. Therefore, we will type y = a * sin(b * x + c) + d into the box. Please note that here a, b, c, d are parameters that are to be determined under fitting. Then we will click "Submit" to enter the function. And you will see there are four boxes to be filled.

Gerneral Fitting XY		
General Fitting XY		
C Linear Regression	upasay suuri misu (* 1 / 2) Bronse, – lucuor folytight.cov	
Non-Linear Regression	Lipital complete	
Non Linear Regression (XY)	Input Your Function: 7	
Multi Regression	ya'alab'ardid	
Pind What Fits		
Curve Fitting	a: 0	
	Bc 0	
	d 0 do x 15 de r	25
	d: 0	
	Submit Fit 2	
	Best FR Curve	
	Residuals	
	Gquared Residuals	
	Prediction of:	
	0	

This is for the initial guess for the function (to know more, you can look for the gradient descent method online, but let's ignore it for now). We can simply type in 1, 1, 1, 1 for all values. After that, we click "Fit" and then check the "Best Fit Curve". You will see the following:

Gerneral Fitting XY	≡	ĺ
General Fitting XY		1
Linear Regression	Upload your lite: + 1 7 Browsen Hourd Opylight.cov	I
Non-Linear Regression	Updat complex	I
Non Linear Regression XY	Input Your Function: 7	I
Multi Regression	ywfaithhedid	I
Pind What Fits		I
Curve Fitting		I
		I
	a 1	
	d 1	l
		l
		ł
	$y = 1.7043 \cdot \sin(0.55727 \cdot x + 4.4002) + 12.337$	I
	Residual sum of squares 1	
	RSS = 0.77323	1

Together with the graph, the tools also tell you different related information, such as the equation of the best fit curve and the residual sum of squares. The residual sum of squares measures how well this curve can approximate the dataset we have. You can use this to compare which model is better for the regression, too. (Please note that a low residual sum of squares does not always imply the model we are using is a good one.)

2.2.2 Residuals

This gives us a set of vertical lines, showing the difference between each of our original data points and the regression curve.



2.2.3 Squared Residuals

This function visualizes the square of residuals as the area of the rectangles. In fact, it is originally a square with side length equal to the residual, just that in the graph, the vertical and horizontal scales are different, and that makes the shape look like a rectangle.



2.2.4 Prediction

By typing any number into the box, we can know the corresponding prediction of the y-coordinate according to our regression curve. You can also see how the regression line is extended, and the intersection with the red vertical line gives you the prediction of the new value.

For example, in our case, if we want to predict the average hours of daylight of March in the next year, we can type 15 into the box and we can see it will be about 12.7 hours.



3 Conclusion

Nonlinear regression is a very powerful tool in mathematical modelling. By considering different nonlinear regression models, we can understand the trends in datasets and make predictions. We hope that this R Shiny General Fitting (XY) tool may help you with your mathematical modelling journey. Good luck!