SOLUTION OF REGRESSION MODELS

NonLinear Regression Model Comparison - with Data linearization vs. without Data linearization

2007 *Jamie Trahan , Autar Kaw*
University of South Florida
United States of America
kaw@eng.usf.edu
http://numericalmethods.eng.usf.edu

Introduction

This worksheet illustrates finding the constants of nonlinear regression models withou linearization. Three common nonlinear models are illustrated -

1) **Exponential:** $y = ae^{bx}$
2) **Power:** $y = ax^b$
3) **Saturation:** $y = \frac{(ax)}{(b+x)}$

where *a* and *b* are constants of the model.

Given *n* data points (x_1,y_1), (x_2, y_2), (x_3,y_3),....,(x_n,y_n), you can best fit one of the nonlinear models to the data. To learn more about nonlinear regression models see the **Nonlinear Regression** model worksheet.

Section 1: Input Data

Below are the input parameters to begin the simulation. This is the only section that re user input. The user can change the highlighted values and Mathcad will return the co of the nonlinear regression model that are derived with data linearization and without linearization.

NOTE: Before evaluating the worksheet, the user must enter initial guesses of the con of the model *a* and *b* for the nonlinear regression without data linearization procedure reasonable initial guesses, use the solution from the **Nonlinear model with data linearization** worksheet. For convergence use initial guesses for *a* and *b* close to the values of *a* and *b* obtained by using data linearization.

NOTE: The origin has been set to 1 to redefine that starting index of all arrays. The u SHOULD NOT change this value.

ORIGIN := 1

Input Parameters:

- Array of *x* values, **X**

$$X := \begin{pmatrix} 10 \\ 16 \\ 25 \\ 40 \\ 60 \end{pmatrix}$$

- Array of *y* values, **Y**

$$Y := \begin{pmatrix} 94 \\ 118 \\ 147 \\ 180 \\ 230 \end{pmatrix}$$

- For *exponential* model call model type to be **"Exponential"**
  For *power* model assign the model type variable as **"Power"**
  For *saturation growth* model, assign the model type variable to be **"Growth"**

ModelType := "Exponential"

- Insert your initial guesses for *a* and *b* here. Reasonable initial guesses for *a* and *l* be obtained from data linearization models.

  Initial Guess values for *a*

  Ainit := 86.9

  Initial Guess values for *b*

  Binit := 0.0159

NOTE: In Section 3, *Solve Block* is used to find the constants of the regression mod without data linearization. You may need to use a different method to solve if you aı unable to get a solution. Change the algorithm by right clicking on the *Find* commaı and checking a different nonlinear method in the popup menu.

- Number of data points, *n*

  n := rows(X)

  n = 5

Section 2: Nonlinear model with data linearization procedure

In the procedure below, the data is first linearized so that least squares regression met for a linear model can be used. Once the coefficients of the linear model are determin the constants of the nonlinear regression model *a* and *b* can be calculated.

Linearizing the data is a useful technique to estimate the parameters of a nonlinear mo because it does not require iterative methods to solve for the model constants. Note th data linearization is only done for mathematical convenience. For details, see the **Nonlinear Regression** worksheet.

```
datalinearized(x, y, n) := | z ← 0
                           | q ← 0
                           | sumq ← 0
                           | sumz ← 0
                           | sumqz ← 0
                           | sumqq ← 0
                           | for  i ∈ 1 .. n
                           |    if  ModelType = "Exponential"
                           |       z_i ← ln(y_i)
                           |       q_i ← x_i
                           |    if  ModelType = "Power"
                           |       z_i ← ln(y_i)
                           |       q_i ← ln(x_i)
                           |    if  ModelType = "Growth"
                           |       z_i ← 1/y_i
                           |       q_i ← 1/x_i
                           |    sumq ← sumq + q_i
                           |    sumz ← sumz + z_i
                           |    sumqz ← sumqz + q_i·z_i
                           |    sumqq ← sumqq + (q_i)^2
                           | a1 ← n·sumqz − sumq·sumz
```

Linearizing the data and making substitutions

Calculating the constants a_0 and a_1 of linear model $z = a_0 + a_1 q$

$$a1 \leftarrow \frac{\quad}{n \cdot sumqq - sumq^2}$$

$$a0 \leftarrow \frac{sumz}{n} - a1 \cdot \frac{sumq}{n}$$

if ModelType = "Exponential"

$$a \leftarrow \exp(a0)$$
$$b \leftarrow a1$$

if ModelType = "Power"

$$a \leftarrow \exp(a0)$$
$$b \leftarrow a1$$

if ModelType = "Growth"

$$a \leftarrow \frac{1}{a0}$$
$$b \leftarrow \frac{a1}{a0}$$

$$\begin{pmatrix} a \\ b \end{pmatrix}$$

Using substitutions of linear model constants to find the constants of the original nonlinear model, *a* and *b*.

Calling the data linearization procedure for the input data set and returning the consta the nonlinear regression model *a* and *b*.

$$a := datalinearized(X, Y, n)_1$$

$$b := datalinearized(X, Y, n)_2$$

$$g(x) := \begin{cases} a \cdot \exp(b \cdot x) & \text{if ModelType = "Exponential"} \\ a \cdot x^b & \text{if ModelType = "Power"} \\ \dfrac{a \cdot x}{b + x} & \text{if ModelType = "Growth"} \end{cases}$$

The constants of the ModelType = "Exponential" nonlinear model found with data linearization are:

$a = 87.8045128$

$b = 0.0169529$

Section 3: Nonlinear model without data linearization procedure

In the following procedure, the constants of the nonlinear regression model *a* and *b* ar found without linearizing the data. This requires the use of a **Solve Block** which utiliz numerical techniques to converge to a real solution. In this section, the initial guess in from Section 1, *Ainit* and *Binit*, are used as the starting values.

Assigning the proper regression model:

$$f(x,aa,bb) := \begin{vmatrix} aa \cdot \exp(bb \cdot x) \ \text{ if } \ ModelType = \text{"Exponential"} \\ aa \cdot x^{bb} \ \text{ if } \ ModelType = \text{"Power"} \\ \dfrac{aa \cdot x}{bb + x} \ \text{ if } \ ModelType = \text{"Growth"} \end{vmatrix}$$

Calculating the sum of the square of the residuals, *Sr*:

$$Sr(x,aa,bb,y,n) := \begin{vmatrix} Sr \leftarrow 0 \\ \text{for } \ i \in 1..n \\ \quad Sr \leftarrow Sr + \left(y_i - f\left(x_i, aa, bb\right)\right)^2 \\ Sr \end{vmatrix}$$

Differentiating S_r with respect to the constants of the model *a* and *b* to setup two simultaneous nonlinear equations and two unknowns:

$$q(aa,bb) := \frac{d}{daa} Sr(X,aa,bb,Y,n)$$

$$r(aa,bb) := \frac{d}{dbb} Sr(X,aa,bb,Y,n)$$

Solving the two simultaneous nonlinear equations. We are using a *Solve Block* sinc we are looking for real solutions.

NOTE: If the solution is not close to the values obtained by data linearization, char the algorithm that is used by right clicking on the *Find* command and checking a different nonlinear method in the popup menu.

Guess

$aa := Ainit$

$bb := Binit$

Given

$q(aa, bb) = 0$

$r(aa, bb) = 0$

constants_of_model := Find(aa, bb)

$aa := constants\_of\_model_1$

$bb := constants\_of\_model_2$

The constants of the ModelType = "Exponential" model without data linearization are

aa = 92.3399879

bb = 0.0155851

Data Table:

$$data\_compared := \begin{pmatrix} a & b \\ aa & bb \end{pmatrix}$$

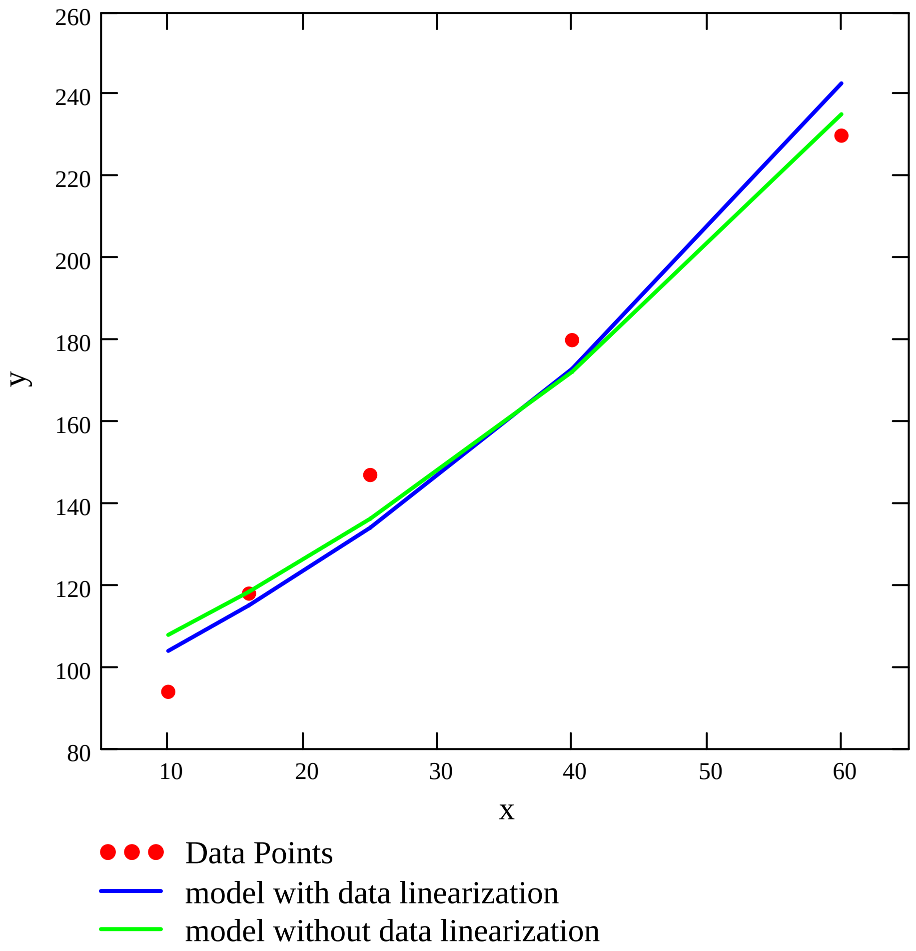
data_compared =

	a	b	
	1	2	
1	87.8045128	0.0169529	**data linearization**
2	92.3399879	0.0155851	**without data linearization**

Section 4: Plot of nonlinear model with data linearization and without data linearization

Plotting the observed values and both predicted curves.

Regression model with data linearization and data without linearization, y vs x



References

[1]Autar Kaw, *Holistic Numerical Methods Institute, http://numericalmethods.eng.usf.edu*, See
Nonlinear Regression

Conclusion

Mathcad helped us to compare a nonlinear regression model that was determined by linearizing the data to one that was found without linearizing the data.

Question 1: A functional relationship, $\rho = k_1 exp(-k_2 h)$ exists between the mass dens of air and the altitude above sea level , h.

$$i := 1 .. 6$$

Altitude above sea level (*km*) $h_i :=$	Mass Density of air (kg/m^3) $\rho_i :=$
0.32	1.15
0.5	1.12
0.64	1.10
1.28	1.05
1.35	1.03
1.60	0.95

a) Find the constants of the model, k_1 and k_2 using data linearization.

b) Find the constants of the model, k_1 and k_2 without data linearization.

c) Compare the two models from (a) and (b) on a plot.

d) Find the sum of the square of the residuals for both models using the original mode $\rho = k_1 exp(-k_2 h)$. Which one is smaller? Is that what you expected?

Question 2: Theoretical considerations assume that the rate of flow, Q from a fire hos proportional to some power of the nozzle pressure, p as given by $Q = ap^b$. Determine whether this conjecture is true.

$i := 1 .. 7$

Flow rate (*gallons/min*)	Pressure (*psi*)
$Q_i :=$	$p_i :=$
91	9
120	15
127	23
190	40
240	61
310	72
409	90

a) Find the constants, *a* and *b* of the model using data linearization.
b) Find the constants, *a* and *b* of the model without data linearization.
c) Compare the two models on a plot.
d) Find the sum of the square of the residuals for both the models using the original m
$Q = ap^b$. Which one is smaller? Is that what you expected?

---

equires
onstants
data

nstants
. For

user

nts of

ity, ρ

el

e is