

Nonlinear Regression Model Comparison - with Data linearization vs without Data linearization

*Autar Kaw, Jamie Trahan
University of South Florida
United States of America
kaw@eng.usf.edu*

Introduction

This worksheet illustrates comparing a nonlinear regression model determined with data linearization versus a model without data linearization. The comparison is illustrated for three common nonlinear models.

Exponential: $y = ae^{bx}$

Power: $y = ax^b$

Saturation Growth: $y = \frac{ax}{b+x}$

where a and b are the constants of the above regression models.

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

```
ClearAll;  
Off[General::spell]
```

```
Clear[a]  
Clear[b]
```

Section 1: Input Data

Below are the input parameters to begin the simulation. This is the only section that requires user input. The user can specify the nonlinear model that is used with the *modeltype* variable. Once the values are entered, *Mathematica* will return the constants of the nonlinear regression model that are derived with data linearization and without using data linearization.

NOTE: Before evaluating the worksheet, the user must enter initial guesses of the constants of the model a and b for the nonlinear regression without data linearization procedure. For 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.

Input Parameters:

- Number of data points, n

```
n = 5
```

```
5
```

- Array of x values, X

```
x = {10, 16, 25, 40, 60}
```

```
{10, 16, 25, 40, 60}
```

- Array of y values, Y

```
y = {94, 118, 147, 180, 230}
```

```
{94, 118, 147, 180, 230}
```

- 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"
```

```
Exponential
```

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

Initial Guess values for a

```
Ainit = 86.9
```

```
86.9
```

Initial Guess values for b

```
Binit = 0.0159
```

```
0.0159
```

Section 2: Nonlinear model with data linearization procedure

In the procedure below, the data is first linearized so that least squares linear regression method for a linear model can be used. Once the coefficients of the linear model are determined, 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 model because it does not require iterative methods to solve for the model constants. Note that data linearization is only done for mathematical convenience. For details, see the Nonlinear Regression worksheet.

```

datalinear[x_, y_, n_] := Module[{},
  z = Array[0, {1, n}];
  q = Array[0, {1, n}];
  sumq = 0;
  sumz = 0;
  sumqz = 0;
  sumqq = 0;
  (* Data Linearization Step *)
  For[i = 1, i ≤ n, i++,
    If[modeltype == "Exponential", z[[1, i]] = N[Log[y[[i]]]];
      q[[1, i]] = x[[i]];
    If[modeltype == "Power", z[[1, i]] = N[Log[y[[i]]]];
      q[[1, i]] = N[Log[x[[i]]]];
    If[modeltype == "Growth", z[[1, i]] = N[1/y[[i]]];
      q[[1, i]] = 1/x[[i]];
    (* Calculating the constants of the Linear model, z= a0+a1q *)
    sumq = sumq + q[[1, i]];
    sumz = sumz + z[[1, i]];
    sumqz = sumqz + q[[1, i]] * z[[1, i]];
    sumqq = sumqq + (q[[1, i]])^2];
  a1 = N[ $\frac{n * sumqz - sumq * sumz}{n * sumqq - sumq^2}$ ];
  a0 = N[ $\frac{sumz}{n} - a1 * \frac{sumq}{n}$ ];
  (* Calculating the constants
    of the original Nonlinear Regression model, a and b *)
  If[modeltype == "Exponential",
    a = E^a0;
    b = a1];
  If[modeltype == "Power",
    a = E^a0;
    b = a1];
  If[modeltype == "Growth",
    a = 1 / a0;
    b = a1 / a0];
  a; b]

```

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

```

datalinear[X, Y, n];
If[modeltype == "Exponential", g[x_] := N[a * E^(b * x)]];
If[modeltype == "Power", g[x_] := a * x^b];
If[modeltype == "Growth", g[x_] := (a * x) / (b + x)];
Print["The constants of the ", modeltype,
  " model found with data linearization are a = ", a, " and b = ", b]

```

```

The constants of the Exponential
model found with data linearization are a = 87.8045 and 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 are found without linearizing the data. This requires the *FindRoot* command which utilizes numerical techniques to converge to a real solution. In this section, the initial guess inputs from Section 1, *Ainit* and *Binit* are used as the starting values.

```
Clear[aa];
Clear[bb];
Clear[Sr];
```

Assigning the proper regression model

```
If[modeltype == "Exponential", f[x_] = aa * E^(bb * x)];
If[modeltype == "Power", f[x_] = aa * x^bb];
If[modeltype == "Growth", f[x_] = (aa * x) / (bb + x)];
```

Calculating the sum of the square of the residuals

```
Sr = 0;
For[i = 1, i ≤ n, i++, Sr = Sr + ((Y[[i]] - f[X[[i]]])^2)];

Sr
```

$$(94 - aa e^{10 bb})^2 + (118 - aa e^{16 bb})^2 + (147 - aa e^{25 bb})^2 + (180 - aa e^{40 bb})^2 + (230 - aa e^{60 bb})^2$$

Differentiating the sum of the square of the residuals with respect to the constants of the model, a and b , to setup two simultaneous nonlinear equations and two unknowns.

```
eqn1 = Expand[D[Sr, aa] == 0];
eqn2 = Expand[D[Sr, bb] == 0];
```

Solving the two simultaneous nonlinear equations. We are using *FindRoot* since we are looking for real solutions.

```
soln = FindRoot[{eqn1, eqn2}, {{aa, Ainit}, {bb, Binit}}, MaxIterations → 1000];
aa /. soln[[1]];
aa = %;
bb /. soln[[2]];
bb = %;
Print["The constants of the ", modeltype,
      " model without data linearization are a = ", aa, " and b = ", bb]
```

The constants of the Exponential model without data linearization are a = 92.34 and b = 0.0155851

■ Comparison Table

The following table allows you to compare the coefficient values calculated between the two methods.

```
TableForm[{{a, b}, {aa, bb}}, TableDepth → 2,  
TableHeadings →  
{{"with data linearization", "without data linearization"}, {"a", "b"}}]
```

	a	b
with data linearization	87.8045	0.0169529
without data linearization	92.34	0.0155851

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

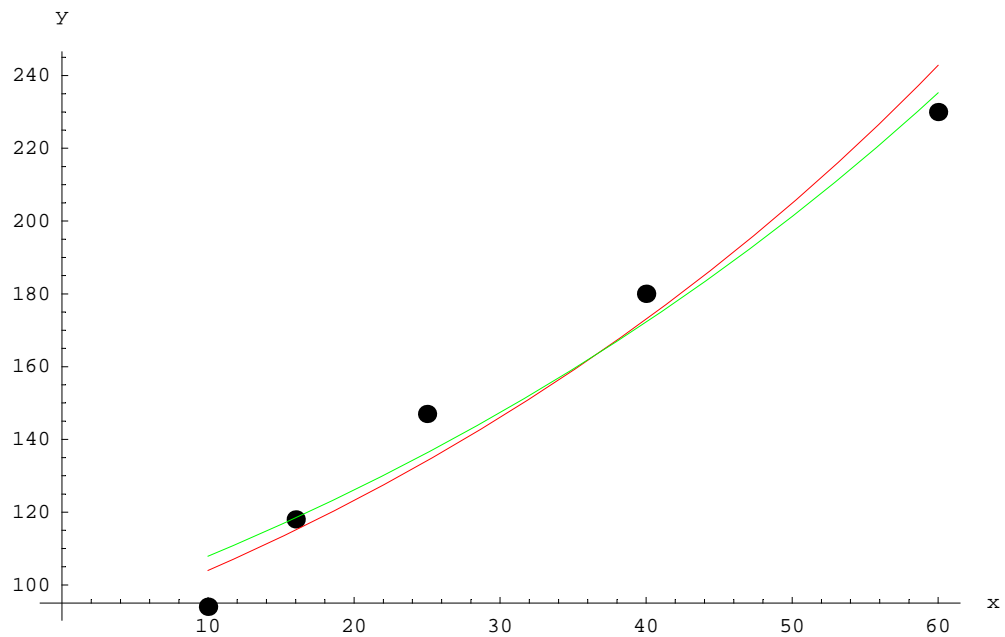
Plotting the observed values and both predicted curves.

```

Observed = Table[{X[[i]], Y[[i]]}, {i, 1, n}];
Predicted1 = g[x];
Predicted2 = f[x];
titl = Print[modeltype,
  " Regression model with data linearization and without linearization"];
points = ListPlot[Observed, PlotStyle -> PointSize[.02], DisplayFunction -> Identity];
lin = Plot[{Predicted1, Predicted2}, {x, Min[X], Max[X]},
  PlotStyle -> {{RGBColor[1, 0, 0]}, {RGBColor[0, 1, 0]}}, DisplayFunction -> Identity];
Show[points, lin, AxesLabel -> {x, y}, DisplayFunction -> $DisplayFunction]
Print["model with data linearization = red"]
Print["model without data linearization = green"]

```

Exponential Regression model with data linearization and without linearization



- Graphics -

model with data linearization = red

model without data linearization = green

Conclusion

Mathematica 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, $\rho = k_1 e^{-k_2 h}$ exists between the mass density, ρ of air and the altitude above sea level, h .

Altitude above sea level, h (km)	0.32	0.5	0.64	1.28	1.35	1.60
Mass Density of air, ρ (kg / m^3)	1.15	1.12	1.10	1.05	1.03	0.95

- Find the constants, k_1 and k_2 of the model using data linearization
- Find the constants, k_1 and k_2 of the model without data linearization
- Compare the two models from (a) and (b) on a plot.
- Find the sum of the square of the residuals for both the models. Which one is smaller? Is that what you expected?

Question 2: Theoretical considerations assume that the rate of flow from a fire hose is proportional to some power of the nozzle pressure. Determine whether this conjecture is true.

Flow rate (gallons / min) F	91	120	127	190	240	310	409
Pressure (psi) p	9	15	23	40	61	72	90

- Find the constants, a and b , of the model using data linearization
- Find the constants, a and b , of the model without data linearization
- Compare the two models on a plot
- Find the sum of the square of the residuals for both the models. Which one is smaller? Is that what you expected?

References

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