SOLUTION OF REGRESSION MODELS

NonLinear Regression with Data Linearization

2007 *Fabian Farelo , Autar Kaw, Jamie Trahan*
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 with data linearization. Three common nonlinear models are illustrated - 1) Exponential 2) Power 3) Saturation Growth.

Given n data points (x_1, y_1), (x_2, y_2), (x_3, y_3), ..., (x_n,y_n), best fit one of the following models to the data

Exponential: $y = ae^{bx}$ (1.1)

Power: $y = ax^b$ (1.2)

Saturation: $y = \dfrac{ax}{b+x}$ (1.3)

a and b are constants of the regression model.

Section 1: Input

The following data table is required to begin the simulation. The user can change the values by clicking on the table and entering the *x*-coordinate in the first column and the *y*-coordinate in the second column. The user can choose the nonlinear model that is desired with the *model_type* variable.
Note: the origin has been set to one to redefine the starting index of all arrays. The user SHOULD NOT change this value. ORIGIN := 1

data :=

	1	2
1	0.5000	1.0000
2	1.0000	0.8920
3	3.0000	0.7080
4	5.0000	0.5620
5	7.0000	0.4470
6	9.0000	0.3550

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"

model_type := "exponential"

Section 2: Extracting the Data

The following values are extracted from the input data table allowing us to use the data in the procedures below.

$x := data^{\langle 1 \rangle}$

$y := data^{\langle 2 \rangle}$

$n := rows(x)$

$n = 6$

Section 3: Finding the constants of the Model

The following manipulation must first be made to each model in order to linearize the data:

Exponential: $\ln(y) = \ln(a) + bx$ (*taking the natural log of both sides*) (2.1)
Power: $\ln(y) = \ln(a) + b\ln(x)$ (*taking the natural log of both sides*) (2.2)
Growth: $\frac{1}{y} = \frac{b}{a}\frac{1}{x} + \frac{1}{a}$ (*taking the reciprocal of both sides then rearranging*) (2.3)

Once the data is linearized, substitutions are made so as to apply a direct solution approach using least squares regression method for a straight line. In this section, a procedure is written for each model. Once a specific nonlinear model is called for with the *model_type* variable, the proper procedure will calculate the coefficients for that model.
Note: See Nonlinear Regression model notes for explanation of the model.

Exponential model procedure:

After data linearization, the following substitutions are made

let $z = \ln(y)$

$a_0 = \ln(a)$, implying $a = e^{a0}$

$a_1 = b$

The data z versus x now takes the form of a linear model:

$$z = a_0 + a_1 x \qquad (2.4)$$

In the *exponential_model* procedure, least squares linear regression method is used to solve for the a_0 and a_1 coefficients which are then used to determine the original constants of the exponential model, a and b, where $y = ae^{bx}$.

```
exponential_model(x, y, n) := | z ← 0
                              | sumz ← 0
                              | sumx ← 0
                              | sumxz ← 0
                              | sumxx ← 0
                              | for i ∈ 1 .. n
                              |   | z_i ← ln(y_i)
                              |   | sumz ← sumz + z_i
                              |   | sumx ← sumx + x_i
                              |   | sumxz ← sumxz + x_i·z_i
                              |   | sumxx ← sumxx + (x_i)^2
                              | a1 ← (n·sumxz − sumx·sumz) / ((n·sumxx) − sumx^2)
                              | a0 ← (sumz / n) − a1·(sumx / n)
                              | a ← exp(a0)
                              | b ← a1
                              | (a, b)
```

Power model procedure:

After data linearization, the following substitutions are made

let $z = \ln(y)$

$w = \ln(x)$

$a_0 = \ln(a)$, implying $a = e^{a0}$

$a_1 = b$

The data z versus w now takes the form of a linear model:

$$z = a_0 + a_1 w \qquad (2.5)$$

In the *power_model* procedure, least squares linear regression method is used to solve for the a_0 and a_1 coefficients which are then used to determine the original constants of the power model, a and b, where $y = ax^b$.

```
power_model(x, y, n) := | z ← 0
                        | w ← 0
                        | sumz ← 0
                        | sumw ← 0
                        | sumwz ← 0
                        | sumww ← 0
                        | for  i ∈ 1 .. n
                        |   | z_i ← ln(y_i)
                        |   | w_i ← ln(x_i)
                        |   | sumw ← sumw + w_i
                        |   | sumz ← sumz + z_i
                        |   | sumwz ← sumwz + w_i·z_i
                        |   | sumww ← sumww + (w_i)^2
                        | a1 ← (n·sumwz − sumw·sumz) / (n·sumww − sumw^2)
                        | a0 ← sumz/n − a1·sumw/n
                        | a ← exp(a0)
                        | b ← a1
                        | (a)
                        | (b)
```

Saturation Growth model procedure:

After data linearization, the following substitutions are made

let $z = 1/y$

$q = 1/x$

$a_0 = 1/a$, implying $a = 1/a_0$

$a_1 = b/a$ implying $b = a_1/a_0$

The data z versus q now takes the form of a linear model:

$$z = a_0 + a_1 q \qquad (2.6)$$

In the saturation *growth_model* procedure, least squares linear regression method is used to solve for the a_0 and a_1 coefficients which are then used to determine the original constants of the growth model, a and b, where $y = (ax)/(b+x)$.

```
growth_model(x, y, n) := | z ← 0
                         | q ← 0
                         | sumq ← 0
                         | sumz ← 0
                         | sumqz ← 0
                         | sumqq ← 0
                         | for i ∈ 1 .. n
                         |   | 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) / (n·sumqq − sumq^2)
                         | a0 ← sumz/n − a1·sumq/n
                         | a ← 1/a0
                         | b ← a1/a0
                         | (a)
                         | (b)
```

The constants of the desired nonlinear model are returned by the following procedure:

$$\text{constants\_of\_model} := \begin{vmatrix} c \leftarrow \text{exponential\_model}(x,y,n) \ \text{ if } \ \text{model\_type} = \text{"exponential"} \\ c \leftarrow \text{power\_model}(x,y,n) \ \text{ if } \ \text{model\_type} = \text{"power"} \\ c \leftarrow \text{growth\_model}(x,y,n) \ \text{ if } \ \text{model\_type} = \text{"growth"} \\ c \end{vmatrix}$$

$a := \text{constants\_of\_model}_1$ $b := \text{constants\_of\_model}_2$

$a = 1.026$ $b = -0.119$

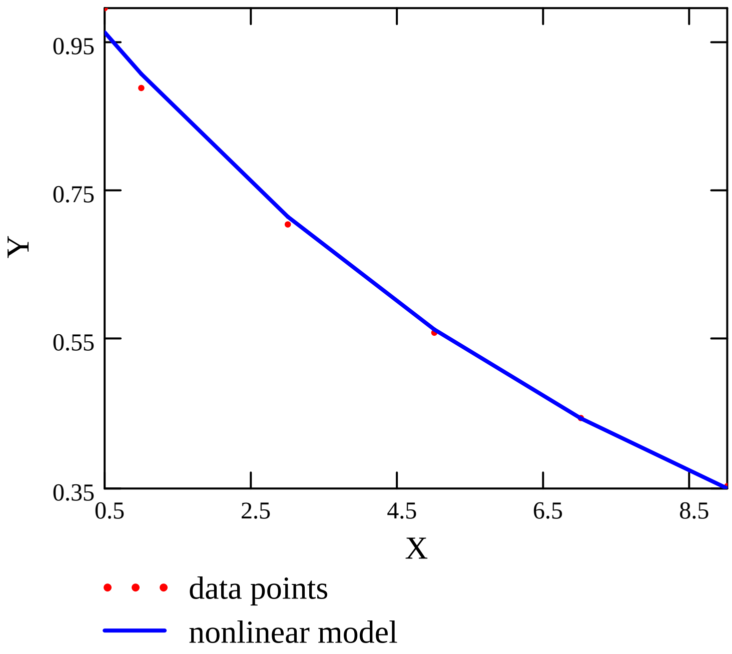
Below, the nonlinear model is plotted versus the data set.

$$f(x) := \begin{vmatrix} a \cdot \exp(b \cdot x) \ \text{ if } \ \text{model\_type} = \text{"exponential"} \\ a \cdot x^b \ \text{ if } \ \text{model\_type} = \text{"power"} \\ \dfrac{a \cdot x}{b + x} \ \text{ if } \ \text{model\_type} = \text{"growth"} \end{vmatrix}$$

Recall that the model type is:

model_type = "exponential"

Nonlinear regression model with data linearization, *y* vs *x*



References

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

Conclusion

Mathcad helped us apply our knowledge of linear regression method to regress a given data set to a nonlinear model.

Question 1: You are working for Valdez SpillProof Oil Company as a petroleum engineer. Your boss is asking you to estimate the future life of an oil well. The analysis used in the industry is called the decline curve analysis where the barrels of oil produced per unit time are plotted against time, and the curve is extrapolated. One of the standard curves used is harmonic decline model, that is

$$q = \frac{b}{1 + at}$$

where q is the rate of production and t is the time, b and a are the constants of the regression model.

Time (t), Month	1	5	11	16	18
Rate of Production (q), Barrels per Day	258	185	109	84	71

a) Find the constants of the regression model. **Hint: You are allowed to linearize the data i possible.**
b) Find the total life of an oil field if 5 barrels per day is considered the production at which t field needs to be abandoned for further production.
c) What does b stand for?

2. It is desired to obtain a functional relationship between the mass density ρ of air and the altitude h above the sea level for the dynamic analysis of bodies moving within earth's atmosphere. Use the approximation

Altitude, (kilometers)	**Mass Density, (kg/m3)**
0.32	1.15
0.64	1.1
1.28	1.05
1.6	0.95

3. It is suspected from theoretical considerations that the rate of flow from a fire hose is proportional to some power of the nozzle pressure. Determine whether the speculation is tru What is the exponent of the data. Assume pressure data is more accurate. You are allowed linearize the data.

Flow rate (gallons/min) F	94	118	147	180	230
Pressure (psi) p	10	16	25	40	60