Nonlinear Regression with Data Linearization

© 2007 Fabian Farelo, Autar Kaw, Jamie Trahan

University of South Florida

United States of America

kaw@eng.usf.edu

Note: This worksheet demonstrates the use of Maple to illustrate the procedure to regress a given data set to a nonlinear model with linearization of data.

▼ 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)
Growth: $y = ax/(b+x)$ (1.3)

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

▼ 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 change the desired nonlinear model below and Maple will calculate the nonlinear regression model using data linearization for the corresponding model that the user chooses.

Input Parameters:
n = Number of data points, *n*
X = Array of *x* values
Y = Array of *y* values

```
>  restart;
   n:=6;
   X:=[0.5,1,3,5,7,9];
   Y:=[1.0,0.8920,0.7080,0.5620,0.4470,0.3550];
```
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:="growth";

$$n := 6$$
$$X := [0.5, 1, 3, 5, 7, 9]$$
$$Y := [1.0, 0.8920, 0.7080, 0.5620, 0.4470, 0.3550]$$
$$model\_type := \text{"growth"}$$

(2.1)

▼ Section 2:Nonlinear Model with Data Linearization Procedures

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: $1/y = (b/a)*(1/x)+(1/a)$ *(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 using data linearization.
Note: See nonlinear model regression notes for explanation of the model.

▼ Exponential model procedure

After data linearization, the following substitutions are made to Equation 2.1
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:
   $z = a_0 + a_1 x$ (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:=proc(X,Y,n)
    local f,x,y,a,b,z,i,sumz,sumx,sumxz,sumxx,a1,a0,x1,y1;
    sumz:=0;
    sumx:=0;
    sumxz:=0;
    sumxx:=0;
```

```
    for i from 1 by 1 to n do
        z[i]:=ln(Y[i]);
        sumz:=sumz+z[i];
        sumx:=sumx+X[i];
        sumxz:=sumxz+X[i]*z[i];
        sumxx:=sumxx+X[i]^2;
    end do;
    a1:=(n*sumxz-sumx*sumz)/((n*sumxx)-(sumx)^2);
    a0:=(sumz/n)-a1*(sumx/n);
    a:=exp(a0);
    b:=a1;
    return(a,b);
    end proc:
```

▼ Power model procedure

After data linearization, the following substitutions are made to Equation 2.2:

let $\quad z = \ln(y)$

$\quad w = \ln(x)$

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

$\quad a_1 = b$

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

$\quad z = a_0 + a_1 w$ $\hspace{4cm}$ (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 exponential model a and b, where $y = ax^b$.

```
>  power_model:=proc(X,Y,n)
   local a0,a1,sum,z,w,sumz,sumw,sumwz,sumww,a,b,i;
   sumz:=0;
   sumw:=0;
   sumwz:=0;
   sumww:=0;
   for i from 1 by 1 to n do
       w[i]:=evalf(ln(X[i]));
       z[i]:=evalf(ln(Y[i]));
       sumw:=sumw+w[i];
       sumz:=sumz+z[i];
```

```
        sumwz:=sumwz+w[i]*z[i];
        sumww:=sumww+(w[i])^2;
    end do:
    a1:=(n*sumwz-(sumw*sumz))/(n*(sumww)-(sumw^2));
    a0:=((sumz/n)-a1*(sumw/n));
    a:=exp(a0);
    b:=a1;
    return(a,b);
    end proc:
```

Saturation Growth model procedure

After data linearization, the following substitutions are made to Equation 2.3:

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:

$$z = a_0 + a_1 q \tag{2.6}$$

In the *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 exponential model a and b, where $y = \dfrac{ax}{b+x}$.

```
> growth_model:=proc(X,Y,n)
    local i,z,q,a0,a1,a,b,sumz,sumq,sumqq,sumqz;
    sumz:=0;
    sumq:=0;
    sumqq:=0;
    sumqz:=0;
    for i from 1 by 1 to n do
        z[i]:=1/Y[i];
        q[i]:=1/X[i];
        sumz:=sumz+z[i];
        sumq:=sumq+q[i];
        sumqz:=sumqz+q[i]*z[i];
        sumqq:=sumqq+q[i]^2;
    end do;
    a1:=(n*sumqz-(sumq*sumz))/(n*(sumqq)-(sumq^2));
    a0:=((sumz/n)-a1*(sumq/n));
```

```
  a:=1/a0;
  b:=a1/a0;
  return(a,b);
  end proc:
```

▼ Section 3: Finding the constants of the model

The proper regression model is assigned below, returning the constants of the model that was chosen with the *model_type* variable.

```
> if (model_type="exponential") then
      exponential_model(X,Y,n):
      a,b:=exponential_model(X,Y,n):
      f:=x->a*exp(b*x):
      print(`The constants of the exponential model are a = `,a,
  ` and b =`,b);
  end if:
  if (model_type="power") then
      power_model(X,Y,n):
      a,b:=power_model(X,Y,n):
      f:=x->a*x^b:
      print(`The constants of the power model are a = `,a,` and
  b =`,b);
  end if:
  if (model_type="growth") then
      growth_model(X,Y,n):
      a,b:=growth_model(X,Y,n):
      f:=x->(a*x)/(b+x):
      print(`The constants of the saturation growth model are a
  = `,a,` and b =`,b);
  end if:
```

The constants of the saturation growth model are a = , 0.4605090102, and b =,   **(4.1)**
    −.3236995442

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

```
> predicted:=f(x);
  observed:=[seq([X[i],Y[i]],i=1..n)];
  ttle:=cat(model_type, ` regression model with data
  linearization, y vs x`):
  plot([predicted,observed],x=X[1]..X[n],style=[line,point],
```

```
title=ttle,legend=["Nonlinear Regression model","Data
Points"]);
```

$$predicted := \frac{0.4605090102\ x}{-0.3236995442 + x}$$

$observed := [[0.5, 1.0], [1, 0.8920], [3, 0.7080], [5, 0.5620], [7, 0.4470], [9, 0.3550]]$

growth regression model with data linearization, y vs x
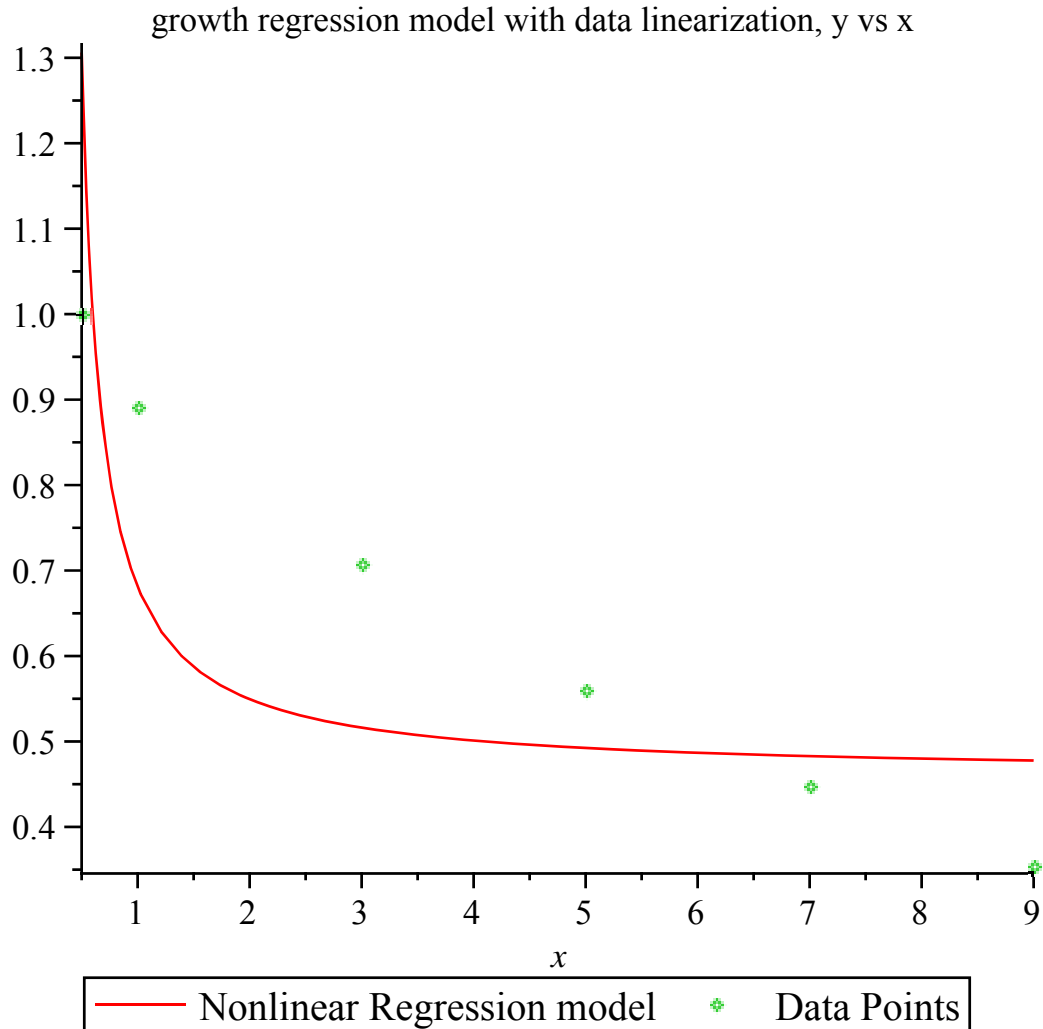
▼ References

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

▼ Conclusion

Maple helped us apply our knowledge of linear regression 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	1\	1\	1\
Rate of Productio\ n (q), Barrels per day	2\	1\	1\	8\	7\

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

Question 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 to fit the data given below by regression analysis to $\rho = k_1 e^{-k2h}$. Find the constants k_1 and k_2. You are allowed to linearize the data.

Altitude($kilometers$\ )	Mass Density($\frac{kg}{m^3}$)
0.32	1.15
0.64	1.10
1.28	1.05
1.60	0.95

Question 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 true. What is the exponent of the data. Assume pressure data is more accurate. You are allowed to linearize the data.

Flow rate ($gallons$/	9\	11\ 4	1\ 8	18\ 47|0	23\ 0

(min)) F					
Pressure (*p.s.i* .) p	1\	16 0	2\	40 5	60