International Journal of Applied Engineering Research ISSN 0973-4562 Volume 12, Number 18 (2017) pp. 7853-7856

Applied Exponential Growth Regression Modeling Using SAS: An Alternative Method of Biostatistics

Wan Muhamad Amir W Ahmad [1], Nor Azlida Aleng[2], Nurfadhlina Abdul Halim[3]
Yosza Dasril[4] and Adam Baharum[5]

[1] *School of Dental Sciences, Health Campus, Universiti Sains Malaysia (USM), 16150 Kubang Kerian, Kelantan, Malaysia.*
[2, 3] *School of Informatics and Applied Mathematics, Universiti Malaysia Terengganu (UMT), Terengganu Malaysia.*
[4] *Department of Industrial Electronics, Faculty of Electronics & Computer Engineering, Universiti Teknikal Malaysia. Melaka (UTeM), Melaka, Malaysia.*
[5] *School of Mathematics Sciences, Universiti Sains Malaysia (USM), Malaysia.*

[2]*Orcid Id: 0000-0003-1111-3388*

Abstract

This paper discussed on how alternative method for exponential growth modeling as a technique for regression analysis through SAS algorithm can be applied to the Biostatistics field. This alternative method is a combination of two major techniques which is include bootstrapping and fuzzy regression for exponential growth model.

Keywords: Bootstrap, exponential growth and fuzzy regression.

INTRODUCTION TO ALGORITM USING SAS LANGUAGE

This research paper provides an illustration and also algorithm for the exponential growth modeling using bacteria growth dataset. For the case of nonlinear regression, we have to transform the equation from non-linear into a linear form. Multiple linear regressions $Y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_k x_k$ are an extension of simple linear regression and had been used widely in medical research. We used this technique in order to get a better result. The random error term is added to make the model probabilistic rather than deterministic. The value of the coefficient β_i determines the contribution of the independent variables x_i, and β_0 is the *y*-intercept (Diem Ngo & La Puente, 2012). A fuzzy regression model can be written as $Y = Z_0 + Z_1 x_1 + Z_2 x_2 + \ldots + Z_k x_k$, here the explanation variables x_i's are assumed to be precise. However, according to the equation above, response variable *Y* is not crisp but is instead fuzzy in nature. That means the parameters are also fuzzy in nature. Our aim is to estimate these parameters. In further discussion, Z_i's are assumes symmetric fuzzy numbers which can be presented by interval. For example, Z_i can be express as fuzzy set given by $Z_i = < a_{1c}, a_{1w} >$ where a_{ic} is centre and a_{iw} is radius or vagueness associated. Fuzzy set above reflects the confidence in the regression coefficients around a_{ic} in terms of symmetric triangular memberships function. Application of this method should be given more attention when the underlying phenomenon is fuzzy which means that the response variable is fuzzy. So, the relationship is also considered to be fuzzy. This $Z_i = < a_{1c}, a_{1w} >$ can be written as $Z_1 = [a_{1L}, a_{1R}]$ with $a_{1L} = a_{1c} - a_{1w}$ and $a_{1R} = a_{1c} - a_{1w}$. In fuzzy regression methodology, parameters are estimated by minimizing total vagueness in the model. $y_j = Z_0 + Z_1 x_{1j} + Z_2 x_{2j} + \ldots + Z_k x_{kj}$.

Using $Z_i = < a_{1c}, a_{1w} >$ we can write $y_j = < a_{0c}, a_{0w} > + < a_{1c}, a_{1w} > x_{1j} + \ldots + < a_{nc}, a_{nw} > x_{nj} = < a_{jc}, a_{jw} >$. Thus this can be written as $y_{jc} = a_{0c} + a_{1c} x_{1j} + \cdots + a_{nc} x_{nj}$

then it can be written straightly as $y_{jw} = a_{0w} + a_{1w}|x_{1j}| + \cdots + a_{nw}|x_{nj}|$. As y_{jw} represent radius and so cannot be negative, therefore on the right-hand side of equation $y_{jw} = a_{0w} + a_{1w}|x_{1j}| + \cdots + a_{nw}|x_{nj}|$, absolute values of x_{ij} are taken. Suppose there m data point, each comprising $a(n+1)-row$ vector.

Then parameters Z_i are estimated by minimizing the quantity, which is total vagueness of the model-data set combination, subject to the constraint that each data point must fall within estimated value of response variable.

This can be visualized as the following linear programming problem, minimized $\sum_{j=1}^{m}\left(a_{0w} + a_{1w}|x_{1j}| + \cdots + a_{nw}|x_{nj}|\right)$ and subject to

$$\left\{\left(a_{0c} + \sum_{i=1}^{n} a_{ic}x_{ij}\right) + \left(a_{0w} + \sum_{i=1}^{n} a_{iw}x_{ij}\right)\right\} \geq Y_j \text{ and}$$

$$\left\{\left(a_{0c} + \sum_{i=1}^{n} a_{ic}x_{ij}\right) - \left(a_{0w} + \sum_{i=1}^{n} a_{iw}x_{ij}\right)\right\} \leq Y_j .$$

and $a_{iw} \geq 0$. Simple procedure is commonly used to solve the linear programming problem. (Kacprzyk and Fedrizzi, 1992). Data of this study is a sample which composed of two variables.

1.1 Algorithm for Exponential Growth

Exponential growth formula is given by $Y = Ae^{bX}$. The first step which is necessary is to transform the growth formula into a linear form. After transforming the formula we obtained the following equation

$$\begin{aligned} \ln Y &= \ln(Ae^{bX}) = \ln(A) + \ln(e^{bX}) \\ &= \ln(A) + bx \qquad \text{(i)} \end{aligned}$$

Exponential decay formula is given by $Y = Ae^{-bX}$. After transforming the formula we obtained the following equation

$$\begin{aligned} \ln Y &= \ln(Ae^{-bX}) = \ln(A) + \ln(e^{-bX}) \\ &= \ln(A) - bx \qquad \text{(ii)} \end{aligned}$$

We can estimate the parameter A and parameter b through the algorithm below:

```
title "Exponential Equation";
ods graphics / imagename = "Exponential Equation";
proc nlin data=boot1 plots=fit;
parameters S0=1 b=0;
model y = S0 * exp(b*x);
ods output EstSummary=summExp;
run;
```

We divide these analyses into Part I and Part II. The aim of the section is to estimate the value of S0 and b. Second Part of this analysis is to use the value of Ln (S0), (it is the ac value) and b (it is the b value). So we put the obtained value into the algorithm in the Part II. Then, the algorithm straightly processes the results.

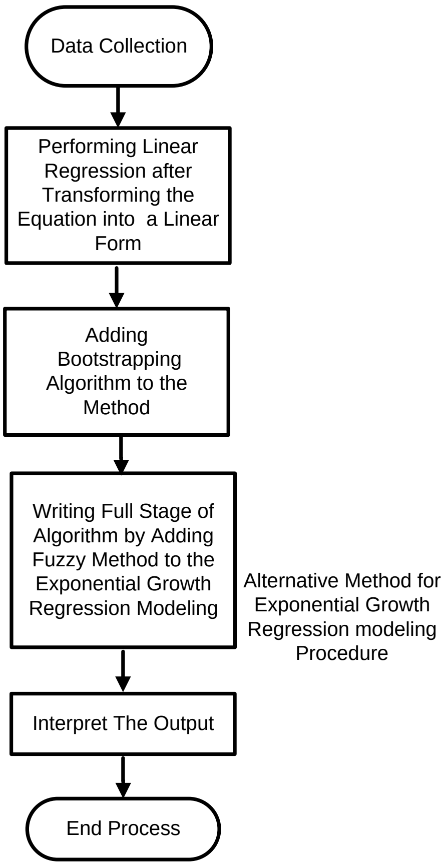

Figure 1: Flow Chart of an Alternative Exponential Modeling

PART I: EXPONENTIAL GROWTH

```
/*Building Basic Algorithm to the Exponential Growth Method*/

data bacteria;
input x y;
datalines;
22      7.33
2       7.33
3       9.86
4       9.68
4       9.68
13      7.32
13      7.32
15      6.81
2       7.33
2       7.33
4       9.68
4       9.68
9       8.52
9       8.52
9       8.52
13      7.32
;
run;

/* PART ONE OF THE ANALYSIS*/
ods rtf file='robdunc0.rtf' style=journal;

title "Performing bootstrap with case resampling";
proc surveyselect data=bacteria out=boot1 method=urs
samprate=1 outhits rep=2;
run;

/* Print Out the Data*/
proc print data= boot1;
run;

title "Exponential Equation";
ods graphics / imagename = "Exponential Equation";
proc nlin data=boot1 plots=fit;
parameters S0=1 b=0;
model y = S0 * exp(b*x);
ods output EstSummary=summExp;
run;
```

PART II: EXPONENTIAL GROWTH

```
/* PART TWO OF THIS ANALYSIS*/
/* Method of fuzzy least squares (FLS) to the above data
*/

proc nlp;
min Y;
decvar ar br ac bc;
bounds ar>=0, br>=0, ac= 9.4201, bc= -0.0154;
lincon ac+2*bc-ar-2*br<=7.33;
lincon ac+2*bc-ar-2*br<=7.33;
lincon ac+3*bc-ar-3*br<=9.86;
lincon ac+4*bc-ar-4*br<=9.68;
lincon ac+4*bc-ar-4*br<=9.68;
lincon ac+13*bc-ar-13*br<=7.32;
lincon ac+13*bc-ar-13*br<=7.32;
lincon ac+15*bc-ar-15*br<=6.81;
lincon ac+2*bc-ar-2*br<=7.33;
lincon ac+2*bc-ar-2*br<=7.33;
lincon ac+4*bc-ar-4*br<=9.68;
lincon ac+4*bc-ar-4*br<=9.68;
lincon ac+9*bc-ar-9*br<=8.52;
```

```
lincon ac+9*bc-ar-9*br<=8.52;
lincon ac+9*bc-ar-9*br<=8.52;
lincon ac+13*bc-ar-13*br<=7.32;
lincon ac+2*bc+ar+2*br>=7.33;
lincon ac+2*bc+ar+2*br>=7.33;
lincon ac+3*bc+ar+3*br>=9.86;
lincon ac+4*bc+ar+4*br>=9.68;
lincon ac+4*bc+ar+4*br>=9.68;
lincon ac+13*bc+ar+13*br>=7.32;
lincon ac+13*bc+ar+13*br>=7.32;
lincon ac+15*bc+ar+15*br>=6.81;
lincon ac+2*bc+ar+2*br>=7.33;
lincon ac+2*bc+ar+2*br>=7.33;
lincon ac+4*bc+ar+4*br>=9.68;
lincon ac+4*bc+ar+4*br>=9.68;
lincon ac+9*bc+ar+9*br>=8.52;
lincon ac+9*bc+ar+9*br>=8.52;
lincon ac+9*bc+ar+9*br>=8.52;
lincon ac+13*bc+ar+13*br>=7.32;
Y=16*ar+108*br;
run;
ods rtf close;
```

Table 1: Parameter coefficient Estimates **(**Exponential Growth)

Optimization Results				
Parameter Estimates				
N	Parameter	Estimate	Gradient Objective Function	Active Bound Constraint
1	Ar	2.010100	16.000000	
2	Br	0.024600	108.000000	
3	Ac	9.420100	0	Equal BC
4	Bc	-0.015400	0	Equal BC

LN (Bacteria Reading) = 9.420100 - 0.015400 x (iii)

Standard Errors (2.01010) (0.02460)

Substituting the values of parameter estimates in model (see Table 1) we obtained the fuzzy least square regression for exponential growth equation. Fuzzy least square regression for exponential decay equation.

Table 2: Parameter coefficient Estimates **(**Exponential Decay)

Optimization Results				
Parameter Estimates				
N	Parameter	Estimate	Gradient Objective Function	Active Bound Constraint
1	Ar	2.010100	16.000000	
2	Br	0.055400	108.000000	
3	Ac	9.420100	0	Equal BC
4	Bc	0.015400	0	Equal BC

Second write the equation as follows:

LN (Bacteria Reading) = 9.420100 + 0.015400 x (iv)

Standard Errors (2.01010) (0.055400)

SUMMARY AND DISCUSSION

This paper gives the explanation for an alternative programming method of bootstrap approach to (exponential growth and decay modeling) nonlinear regression procedure using SAS software. The aim for the algorithm building is to provide the researcher with the alternative programming of a data analysis. This method can be applied for the small sample size data especially where the data is very difficult to collect.

ACKNOWLEDGEMENT:

The authors would like to express their gratitude to Universiti Sains Malaysia for providing the research funding (Grant no.304/PPSG/61313187, School of Dental Sciences, USM).

REFERENCES

[1] Diem Ngo, T.H., La Puente, C.A., (2012). The Steps to Follow in a Multiple Regression Analysis. SAS Global Forum 2012: Statistics and Data Analysis, Paper 333, pp. 1-12.

[2] Kacprzyk, J., Fedrizzi, M., (1992). Fuzzy Regression Analysis, Omnitech Press, Warsaw.