

Estimate Outcome Value of Doubling Cell Growth Using Fuzzy Regression Method

W. M. A. W. Ahmad
Universiti Sains Malaysia
Kubang Kerian, Malaysia
wmamir@usm.my

R. A. A. Rohim
Universiti Sains Malaysia
Kubang Kerian, Malaysia
adawiyah5350@yahoo.com

N. H. Ismail
Universiti Sains Malaysia
Kubang Kerian, Malaysia
noorhuda@usm.my

Abstract—In this paper, the fundamental concepts used in different disciplines are described and these concepts are written into one common theme: assessing crossness agreement of observation. Focus was given on continuous measurements and methodological approaches for expressing these methods mathematically, both for cases with and without reference. The build method was tested and applied to biological sciences data. The objective of the study was to determine the agreement between the two methods on microorganism data sets in order to determine if they can be used interchangeably and to estimate the fuzzy and bootstrap parameters of the models. The value of interclass correlation (ICC) by using a linear equation is 0.79661 (substantial agreement) and the value of ICC using fuzzy regression equation is 0.82692 (almost perfect agreement).

Keywords—agreement; fuzzy regression; bootstrap; ICC

I. INTRODUCTION

In social, behavior, physical, biological and medical sciences reliability and accurate measurements serve as the basis for evaluation. As new concepts, theories and technologies continue to develop, new scales, methods, tests, assays, devices and evaluation instruments become available. In simple, intuitive terms, reliability and accurate measurement may simply mean that the new measurement is the same as the truth or agrees with the truth. However, requiring the new measurement to agree to the truth is often impractical, either because (1) we are willing to accept a measurement up to some tolerable (or acceptable) error or (2) the truth is simply not available to us (either because it is not measurable or because it is only measurable with some degree or error). To deal with issues related to both (1) and (2), a number of concepts, methods and theories have been developed [1]. In this paper, the fundamental concepts used in different disciplines are described and compared and written in one common theme: assessing closeness ‘agreement’ of observations. We focus on methodological approaches for expressing these concepts and methods mathematically, and discuss the data structures for which they are to be used, both for cases with and without a reference (or truth) [2]. These approaches were developed for one or more types of the following data structures: predicted data of bacteria growth and original data of bacteria growth when they are available.

A. Fuzzy Linear Regression

In fuzzy linear regression analysis two factors, the degree of fitness and the fuzziness of data sets are considered [3]. A fuzzy linear regression model is commonly presented as:

$$\tilde{y} = \tilde{f}_l(x) = \tilde{A}_0 + \tilde{A}_1 x_1 = \tilde{A}_x \quad (1)$$

where $x = [1, x_1]^T$ is a crisp vector of independent variables and \tilde{y} is the estimated fuzzy output. $\tilde{A} = [\tilde{A}_0, \tilde{A}_1]$ is a vector of fuzzy parameters of the fuzzy linear regression model. \tilde{A}_j is presented in the form of symmetric triangular fuzzy numbers denoted by $\tilde{A}_j = (\alpha_j, c_j)$, $j = 0, 1, 2, \dots, N$, where its membership function is shown as:

$$\mu_{A_j}(a_j) = \begin{cases} 1 - \frac{a_j - a_j}{c_j} \\ 1 \end{cases}, a_j - c_j \leq a_j \leq a_j + c_j \quad (2)$$

where a_j is the central value of the fuzzy number and c_j is the spread. Therefore the fuzzy linear regression model can be rewritten as:

$$\tilde{y} = (\alpha_0, c_0) + (\alpha_1, c_1)x_1 \quad (3)$$

However, the interaction between variables and higher order terms is not included in the fuzzy linear regression defined in (1). In fact, interaction between variables and higher order terms often exist in physical systems. A simple procedure is commonly used to solve the linear programming problem [4]. Linear regression is used to study the linear relationship between a dependent variable Y and one or more independent variables X. The dependent variable Y must be continuous, while the independent variables may be continuous, binary, or categorical. The initial judgement of a possible relationship between two continuous variables should always be made on the basis of a scatter plot [5].

B. Bacteria Growth (*Streptococcus Sobrinus*)

Bacteria colonize epithelial surfaces in certain parts of the body and there is a constantly on-going competition for space and resources between bacterial species. Bacteria are ubiquitous and the vast majority is harmless or beneficial to the

host. Our understanding with regard to the effect of microbes on human health has gradually developed from pathogens inducing infections to a mutually beneficial interaction with indigenous microorganisms that contribute to normal human physiology and homeostasis [6]. Mutans streptococci are linked with the growth of dental caries in human. These bacteria are the most common pathogens isolated from human dental plaque, and their prevalence has been reported in epidemiological studies [7]. Several methods have been employed for detecting and identifying mutans streptococci, including culturing, direct enzyme tests, enzyme-linked immunosorbent assays, and DNA probes [8]. The mutans group of oral streptococci consist of seven species [9]. Among the group, streptococci mutans and streptococci sobrinus are most frequently isolated from human dental plaque and closely associated with human dental caries [7, 10].

II. MATERIALS AND METHODS

The data were composed of two variables, which are the days of culture (x) and bacterial growth (y). The rate of exponential growth of a bacterial culture is ln γ . The data was available from recent studies for nurturing selected microorganisms (streptococcus sobrinus). The data were recorded in SAS to evaluate microorganism growth before analyzing by using the new proposed statistical methodologies. These approaches were developed for one or more types of the following data structures: predicted data of bacteria growth and original data of bacteria growth where covariates are available for assessing the impact of various factors on agreement measures. Predicting value was done by the use of the linear equation $y=82.9121e^{0.00512x}$.

TABLE I. SAMPLE OF BACTERIA

Days of culture (x)	Original data *(y)	Predicted data *(y)	Exponential growth rate (ln γ)
1	82	83	4.41
2	84	84	4.43
3	86	84	4.45
5	86	85	4.45
9	87	86	4.47
11	87	88	4.47
13	87	89	4.47

* St. Sobrinus

A. The Algorithm of Exponential Calculation

The algorithm is the way of inserting data in SAS programming and the way of calculating the quality of predicted data and original data using interclass correlation (ICC). The name of the dataset is given as test-data. The data consist of two variables, Read1 and Read2, and pid.

```
data test;
/* Read1 is predicted data */
/* Read2 is original data */
input Read1 Read2 pid;
cards;
83      82      1
84      84      2
84      86      3
85      86      4
86      87      5
```

```
88      87      6
89      87      7
;
run;
datatest_data;
set test;
array s(2) Read;
do judge=1 to 2;
y=s(judge);
output;
end;
run;
ods output CovParms=covp;
procmixed data=test_data;
class judge pid;
model y=;
random intercept /subject=pid;
run;
dataicc;
set covp end=last;
retain bvar;
if subject~="" then bvar=estimate;
if last then icc=bvar/(bvar+estimate);
run;
procprint data=icc;
run;
quit;
```

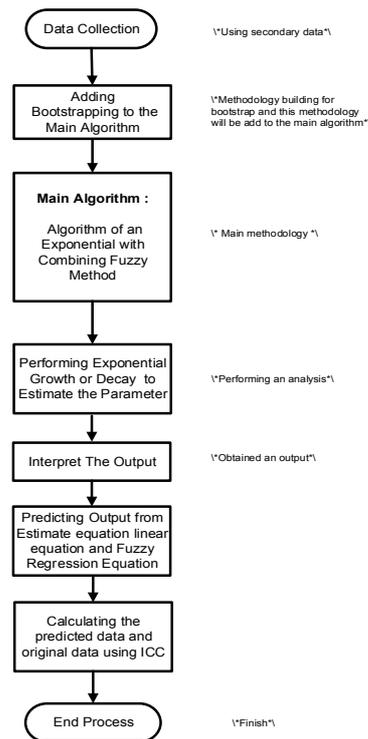


Fig. 1. Algorithm flowchart

B. Assessing Closeness Agreement of Observations Using Linear Equation

The ICC test is usually used in the assessment of agreement between different measurement methods. ICC takes a value

from 0 to 1 (Table II). Table III shows the result of ICC according to the predicted data and original data. The value of ICC is 0.79661 and this is a substantial agreement for the use of linear equation model.

TABLE II. ICC CLASSIFICATION

ICC	Description
<0.00	Poor agreement
0.01-0.20	Slight agreement
0.21-0.40	Fair agreement
0.41-0.60	Moderate agreement
0.61-0.80	Substantial agreement
0.81-1.00	Almost perfect agreement

TABLE III. ICC

Reading	Covariate	Estimate	ICC
1	Intercept	3.3571	
2	Residual	0.8571	0.79661

C. Calculation of Fuzzy Regression For Exponential Growth

The algorithm calculates the fuzzy regression for exponential growth. The complete set of calculations can be visualized by the following linear programming:

```

ods rtf file='abc.rtf' style=journal;
Proc nlp;
min Y;
decvar a0c a0w a1c a1w;
bounds a0w >= 0, a1w >= 0;
lincon a0c+5*a1c-a0w-5*a1w <= 4.45;
lincon a0c+2*a1c-a0w-2*a1w <= 4.43;
lincon a0c+9*a1c-a0w-9*a1w <= 4.47;
lincon a0c+2*a1c-a0w-2*a1w <= 4.43;
lincon a0c+5*a1c-a0w-5*a1w <= 4.45;
lincon a0c+2*a1c-a0w-2*a1w <= 4.43;
lincon a0c+1*a1c-a0w-1*a1w <= 4.41;
lincon a0c+2*a1c-a0w-2*a1w <= 4.43;
lincon a0c+11*a1c-a0w-11*a1w <= 4.47;
lincon a0c+2*a1c-a0w-2*a1w <= 4.43;
lincon a0c+1*a1c-a0w-1*a1w <= 4.41;
lincon a0c+11*a1c-a0w-11*a1w <= 4.47;
lincon a0c+2*a1c-a0w-2*a1w <= 4.43;
lincon a0c+29*a1c-a0w-29*a1w <= 4.55;
lincon a0c+33*a1c-a0w-33*a1w <= 4.6;
lincon a0c+23*a1c-a0w-23*a1w <= 4.54;
lincon a0c+11*a1c-a0w-11*a1w <= 4.47;
lincon a0c+9*a1c-a0w-9*a1w <= 4.47;
lincon a0c+9*a1c-a0w-9*a1w <= 4.47;
lincon a0c+13*a1c-a0w-13*a1w <= 4.47;
lincon a0c+5*a1c-a0w-5*a1w <= 4.45;
lincon a0c+9*a1c-a0w-9*a1w <= 4.47;
lincon a0c+5*a1c-a0w-5*a1w <= 4.45;
lincon a0c+11*a1c-a0w-11*a1w <= 4.47;
lincon a0c+2*a1c-a0w-2*a1w <= 4.43;
lincon a0c+23*a1c-a0w-23*a1w <= 4.54;
lincon a0c+5*a1c+a0w+5*a1w >= 4.45;
lincon a0c+2*a1c+a0w+2*a1w >= 4.43;
lincon a0c+9*a1c+a0w+9*a1w >= 4.47;
lincon a0c+2*a1c+a0w+2*a1w >= 4.43;
lincon a0c+5*a1c+a0w+5*a1w >= 4.45;

```

```

lincon a0c+2*a1c+a0w+2*a1w >= 4.43;
lincon a0c+1*a1c+a0w+1*a1w >= 4.41;
lincon a0c+2*a1c+a0w+2*a1w >= 4.43;
lincon a0c+11*a1c+a0w+11*a1w >= 4.47;
lincon a0c+2*a1c+a0w+2*a1w >= 4.43;
lincon a0c+1*a1c+a0w+1*a1w >= 4.41;
lincon a0c+11*a1c+a0w+11*a1w >= 4.47;
lincon a0c+2*a1c+a0w+2*a1w >= 4.43;
lincon a0c+29*a1c+a0w+29*a1w >= 4.55;
lincon a0c+33*a1c+a0w+33*a1w >= 4.6;
lincon a0c+23*a1c+a0w+23*a1w >= 4.54;
lincon a0c+11*a1c+a0w+11*a1w >= 4.47;
lincon a0c+9*a1c+a0w+9*a1w >= 4.47;
lincon a0c+9*a1c+a0w+9*a1w >= 4.47;
lincon a0c+13*a1c+a0w+13*a1w >= 4.47;
lincon a0c+5*a1c+a0w+5*a1w >= 4.45;
lincon a0c+9*a1c+a0w+9*a1w >= 4.47;
lincon a0c+5*a1c+a0w+5*a1w >= 4.45;
lincon a0c+11*a1c+a0w+11*a1w >= 4.47;
lincon a0c+2*a1c+a0w+2*a1w >= 4.43;
lincon a0c+23*a1c+a0w+23*a1w >= 4.54;
Y=a0w*26+237*a1w;
run;
ods rtf close;

```

D. Parameter Estimation for Fuzzy Regression Exponential Cell Growth

Table IV shows the optimization results for the parameter estimates.

TABLE IV. OPTIMIZATION RESULTS

No	Parameter	Estimate	Gradient Objective Function
1	a0c	4.414107	0
2	a0w	0.009107	26.000000
3	a1c	0.005179	0
4	a1w	0.000179	237.000000

Value of objective function=0.2791071429

From the results shown in Table IV, the fitted model for fuzzy regression is:

$$y = 4.414107 + 0.005179 x_1 \tag{1}$$

E. Assessing Closeness Agreement of Observations Using Fuzzy Regression equation

In Table V we can see the predicted values using fuzzy regression equation from (1):

TABLE V. ORIGINAL VS PREDICTED DATA

Original Data*	Predicted Data*
82	83
84	83
86	84
86	85
87	86
87	87
87	88

* St. Sobrinus (y)

To calculate regression based on fuzzy regression for exponential growth using ICC, the full syntax is:

```

data test;
/* Read1 is predicted data */
/* Read2 is original data */
input Read1 Read2 pid;
cards;
run;

```

TABLE VI. ICC

Reading	Covariate	Estimate	ICC
1	Intercept	3.0714	
2	Residual	0.6429	0.82692

The value of ICC is 0.82692 and this is almost perfect agreement of predicted and original data.

III. SUMMARY AND DISCUSSION

The main objective of this research was to determine the agreement between two methods on microorganism data sets in order to determine if they may be used interchangeably and to estimate the linear and fuzzy regression parameters of the models using interclass correlation. Using linear equation, the value of ICC is 0.79661 which this is substantial agreement for assessing the quality of predicted data, while the value of ICC using fuzzy regression equation is 0.82692 (almost perfect agreement). ICC test is usually used in the assessment of agreement between different measurement methods.

ACKNOWLEDGMENT

Authors would like to express their gratitude to Universiti Sains Malaysia for providing the research funding (Grant no. 1001/PPSG/8012278, School of Dental Sciences, Universiti Sains Malaysia)

REFERENCES

- [1] F. Galton, "Regression Toward Mediocrity in Hereditary Stature", *The Journal of The Anthropological Institute of Great Britain and Ireland*, Vol. 15, pp. 246-263, 1886
- [2] J. M. Bland, D. G. Altman, "Comparing Methods of Measuring: Why Plotting difference against standard method is misleading", *The Lancet*, Vol. 346, pp. 1085-1087, 1995
- [3] H. Tanaka, J. Watada, "Possibilistic Linear Systems and Their Application to the Linear Regression Model", *Fuzzy Sets System*, Vol. 27, pp. 275-289, 1988
- [4] J. Kacprzyk, M. Fedrizzi, *Fuzzy Regression Analysis*, Omnitech Press, Warsaw, 1992
- [5] A. Schneider, G. Homme, M. Bletner, "Linear Regression Analysis", *Deutsches Arzteblatt International*, Vol. 107, No. 44, pp. 776-820, 2010
- [6] S. Rautava, R. Luoto, S. Salminen, E. Isolauri, "Microbiol Contact during Pregnancy, Intestinal Colonization and Human Disease", *Nature Reviews Gastroenterology & Hepatology*, Vol. 9, No. 10, pp. 565-76, 2012
- [7] W. J. Loeshe, "Role of Streptococcus Mutans in Human Dental Decay", *Microbiology Reviews*, Vol. 50, No. 4, pp. 538-80, 1986
- [8] T. Igarashi, A. Yamamoto, N. Goto, "PCR for detection and identification of Streptococcus Sobrinus", *Journal of Medical Microbiology*, Vol. 49, No. 12, pp. 1069-1074, 2000
- [9] R. A. Whaley, D. Beighton, "Current Classification of the Oral Streptococci", *Oral Microbiology and Immunology*, Vol. 13, No. 4, pp. 195-216, 1998
- [10] S. Hamada, H. D. Slade, "Biology, Immunology and Cariogenicity of Streptococcus Mutans", *Microbiology Reviews*, Vo. 44, No. 2, pp. 331-384, 1980