Original Article

Lactobacillus Salivarius Growth: An Algorithm Using Fuzzy Regression Method

Rabiatul Adawiyah Abdul Rohim[1], Puspa Liza Ghazali[2], Kamariah Abdul Jalil[3], Roslida Abdul Razak[4], Juliana Arifin[5], Mohd Sadad Mahmud[6,] and Salman Lambak[7], Eni Noreni Mohamad Zain[8]

[1,3]Faculty of Medicine, Universiti Sultan Zainal Abidin, 20400, Kuala Terengganu, Terengganu, Malaysia
[2,4,5,7,6]Faculty Business and Management, Universiti Sultan Zainal Abidin, 21300 Kuala Nerus, Terengganu, Malaysia
[8]Faculty of Entrepreneurship and Business, Universiti Malaysia Kelantan, 16100 Kota Bharu, Kelantan

Correspondence Email: puspaliza@unisza.edu.my
DOI: 10.47750/pnr.2023.14.S01.13

Abstract

Introduction: The Intraclass Correlation (ICC) is widely used in the assessment of agreement between different measurement methods. Asessing agreement between variables assumes that the variables measure the same construct. In health science research, ICC are widely used to test intrarater and interrater reliability.
Objectives: The objective of the study is to determine the agreement between the two methods on microorganism data set in order to determine if they can be used interchangeable and to estimate the fuzzy regression and bootstrap method parameter of the models.
Methods: This methodology discovers the agreement between predicted data and original data for assessing crossness agreement of two observation. In this paper, linear regression and fuzzy regression have been applied using the SAS algorithm procedure. The ICC was used to measures the agreement and consistency among two quantitative variables.
Results: The value of ICC by using a linear equation is 0.98569 (almost perfect agreement) while the value of ICC using fuzzy regression equation is 0.91212 (almost perfect agreement).
Conclusion: This article also instructs readers on what to look for when encountering ICC in a piece of writing. This study may also motivate researchers to reveal more about their ICC studies, as well as encourage reviewers and editors to demand complete and precise information from authors.

Introduction

Technology in the biological sector is constantly changing at a quick pace, as seen by the regular manufacture of new instruments. One common issue is how the new instrument relates to existing instruments and tests. Instruments provide indirect measurements rather than direct measurements for the eyeA measurement could have been made with a different instrument, with subjective judgement from a human rater, or in a different dataset with a different approach (Streiner and Norman, 2008).

Consistency of measurement can be critical in both research and clinical settings in the health sciences. When measuring equipment provide inconsistent results due to instrumentation flaws, when a single rater produces various measurements of a steady phenomena at different time points, or when different raters produce inconsistent readings among themselves, problems develop (Franco et al., 2014). Inconsistent measurement cannot be overlooked because accurate measurements are frequently required to determine things like appropriate referrals, the extent of disease progression, whether people are amenable to treatment or would be best served by specific types of interventions, and whether people have responded to interventions (Hallgren, 2012). In order to minimise under- and over-diagnosing, measurement consistency is also required, which is difficult to achieve if the assessment bases shift unpredictably. As a result, consistency can be a significant and difficult issue for researchers and physicians. It can be fascinating as well (Haven et al., 2007).

An intra-class correlation coefficient (ICC) and a confidence interval are common ways to evaluate the reliability of measurement processes (Gisev et al., 2013). This method is widely used in biomedical research to determine whether measurements made by raters, labs, technicians or devices are reproducible. The purpose of this study was to observe the agreement between the two methods on microorganism data set in order to determine if they can be interchangeable and to estimate the fuzzy regression and bootstrap method parameters of the models.

Literature Review

Exponential Regression

Exponential growth or decay modeling is a technique for nonlinear regression to study the original nature of the fundamental principles of two variables. In a real world, the exponential growth is often used to model population growth as such cell growth while the exponential decay is often to a model population with declining or decreases in size (Rohim et al., 2020). Exponential growth is often used to model the growth of populations of the organism in a resource-rich environment. Here ‘resource-rich’ means that there is plenty of food and other resources necessary for the population to grow. For example, the initial growth of a cell bacteria in a mouth is often modeled as exponential. The justification for this model is that the rate at which a population of organisms grows should be proportional to the number of organisms, assuming that the organisms reproduce at a constant rate. For example, if you double the size of a population, then this should precisely double the rate at which the population bears offspring, and should, therefore, double the rate at which the size of the population increases. What this means is that the population A of a given organism in a resource-rich environment should satisfy the differential equation $\frac{dA}{dx} = Ax$ value x is some constant that depends on the rate of reproduction. Thus the population grows exponentially $A = A_0 e^{bx}$. This model predicts that the population A will grow indefinitely, which cannot be true in any real situation. Eventually, any population will run out of resources such as food or space to grow. However, the exponential model often gives fairly accurate results in cases where the short-term growth of a population is not inhibited by limited resources (Kreft et al., 1998).

Fuzzy Linear Regression

Tanaka and Watada (1988) were the first to propose fuzzy linear regression technique, which takes into account two factors: degree of fitness and fuzziness of data sets (Mclinden et al, 2011). When dealing with uncertain data, regression is critical in the Fuzzy technique. In this case, a simple regression equation with a single independent fuzzy variable can be utilised.

According to research (Kim et al., 1996), fuzzy regression can be a better replacement for statistical regression when the data is variable and the model specification is insufficient. Bargiela (2007) reported that the regression problem was used to demonstrate the efficiency of an iterative technique for multiple regressions with fuzzy data using gradient descent optimization.

The following is a popular representation of a fuzzy linear regression model:

$$\tilde{y} = \tilde{f}_{Ir}(x) = \tilde{A}_0 + \tilde{A}_1 x_1 = \tilde{A}_X \quad (1)$$

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 $\tilde{A}_J = (\partial_j, c_j), j = 0,1,2,..N$ where its membership function is shown as below:

$$\mu_{AJ}(aj) = \left\{1 - \frac{aj - cj}{cj}\right\}, a_j - c_j \le a_j \le a_j + c_j \qquad (2)$$

Otherwise, 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 shown below.

$$\tilde{y} = (a_0, c_0) + (a_1, c_1)x_1 \qquad (3)$$

The interaction between variables and higher order factors are not taken into account in the fuzzy linear regression defined in (1). Interactions between variables and higher order terms are ubiquitous in physical systems. A simple strategy is commonly used to tackle the linear programming problem (Kreft et al., 1998). Linear regression is used to investigate the linear connection between a dependent variable Y and one or more independent variables X. The dependent variable, Y, must be a continuous variable, while the independent variables can be continuous, binary, or categorical.

Materials and Methods

Data Collection

The data was gathered at the Microbiology Laboratory using several mediums in order to find the best concentration rate. The bacteria's growth rate was determined by monitoring the algal plate at the Microbiology Laboratory, PPSG, USM. The rate of reproduction of Lactobacillus Salivarius bacteria was studied at a concentration of (5x105 CFU/Ml). The bacteria samples were taken on several days (within 7 days). Lactobacillus Salivarius is a type of anaerobic bacterium that does not require oxygen to survive. The dependent variable (Y) in this study was Lactobacillus Salivarius growth rate, while the independent variable (X) was Lactobacillus Salivarius culture days.

Table 1. Sample of Bacteria

Days of Culture (x)	**Original Data *(y)**	**Predicted Data *(y)**	**Exponential Growth Rate (lny)**
1	23	25	3.14
2	23	25	3.14
3	40	44	3.69
5	60	66	4.09
9	80	88	4.38
11	101	111	4.62
13	140	154	4.94

Statistically Analysis

The bootstrap method and fuzzy regression are employed in the main algorithm. ICC is used to calculate the expected and original data. The exponential equation anticipated the data, which was subsequently recorded in SAS programming. The analysis flowchart is shown in Figure 1.

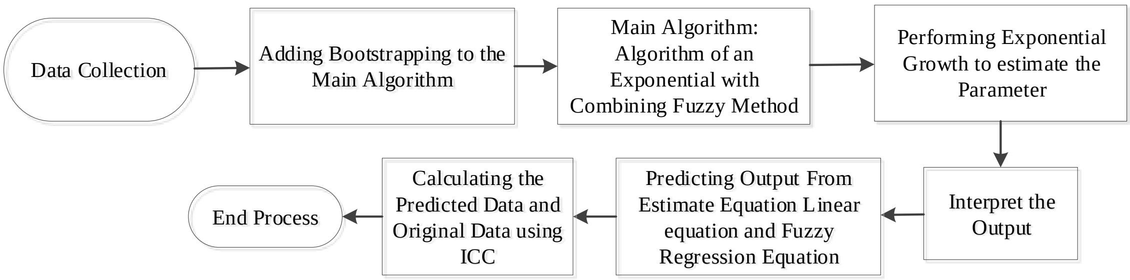

Figure 1: Flowchart of Analysis

Intra-class Correlation

Intraclass correlation (ICC) is a popular measure of agreement for continuous outcomes. Like the Pearson correlation, the ICC requires a linear relationship between the variables. However, it differs from the Pearson correlation in one key respect; the ICC also takes into account differences in the means of the measures being considered. In addition, the ICC can be applied to situations where there are three or more separate raters (Lu N, 2014)

The ICC test is commonly used to determine how well different measuring methods agree. First, we look to see if any information regarding the ICC form was reported and if the correct ICC form was used. The ICC assigns a rating ranging from 0 (no agreement) to 1 (perfect agreement). The ICC classification is as follows: (2016, Koo and Li)

Table 2: Classification of Intraclass Correlation Coefficient

ICC	Explanation
Less than 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

Calculation of Fuzzy Least Squares (FLS) for Exponential Growth

The Fuzzy Least Squares (FLS) for exponential growth is calculated using the algorithm below. The following programming can be used to visualise the entire set of calculations.

```
Proc nlp;
min Y;
decvar a0c a0w a1c a1w;
bounds a0w>=0, a1w>=0;

lincon  a0c+2*a1c-a0w-2*a1w<=3.14;
lincon  a0c+1*a1c-a0w-1*a1w<=3.14;
lincon  a0c+3*a1c-a0w-3*a1w<=3.69;
lincon  a0c+1*a1c-a0w-1*a1w<=3.14;
lincon  a0c+2*a1c-a0w-2*a1w<=3.14;
lincon  a0c+1*a1c-a0w-1*a1w<=3.14;
lincon  a0c+1*a1c-a0w-1*a1w<=3.14;
lincon  a0c+1*a1c-a0w-1*a1w<=3.14;
lincon  a0c+5*a1c-a0w-5*a1w<=4.09;
lincon  a0c+2*a1c-a0w-2*a1w<=3.14;
lincon  a0c+1*a1c-a0w-1*a1w<=3.14;
lincon  a0c+5*a1c-a0w-5*a1w<=4.09;
lincon  a0c+1*a1c-a0w-1*a1w<=3.14;
lincon  a0c+13*a1c-a0w-13*a1w<=4.94;
lincon  a0c+13*a1c-a0w-13*a1w<=4.94;
lincon  a0c+11*a1c-a0w-11*a1w<=4.62;
lincon  a0c+3*a1c-a0w-3*a1w<=3.69;
```

```
lincon  a0c+3*a1c-a0w-3*a1w<=3.69;
lincon  a0c+3*a1c-a0w-3*a1w<=3.69;
lincon  a0c+5*a1c-a0w-5*a1w<=4.09;
lincon  a0c+2*a1c-a0w-2*a1w<=3.14;
lincon  a0c+3*a1c-a0w-3*a1w<=3.69;
lincon  a0c+2*a1c-a0w-2*a1w<=3.14;
lincon  a0c+5*a1c-a0w-5*a1w<=4.09;
lincon  a0c+1*a1c-a0w-1*a1w<=3.14;
lincon  a0c+9*a1c-a0w-9*a1w<=4.38;
lincon  a0c+1*a1c-a0w-1*a1w<=3.14;
lincon  a0c+9*a1c-a0w-9*a1w<=4.09;
lincon  a0c+13*a1c-a0w-13*a1w<=4.94;
lincon  a0c+2*a1c-a0w-2*a1w<=3.14;
lincon  a0c+11*a1c-a0w-11*a1w<=4.62;
lincon  a0c+2*a1c-a0w-2*a1w<=3.14;
lincon  a0c+1*a1c-a0w-1*a1w<=3.14;
lincon  a0c+5*a1c-a0w-5*a1w<=4.09;
lincon  a0c+9*a1c-a0w-9*a1w<=4.38;
lincon  a0c+2*a1c+a0w+2*a1w>=3.14;
lincon  a0c+1*a1c+a0w+1*a1w>=3.14;
lincon  a0c+3*a1c+a0w+3*a1w>=3.69;
lincon  a0c+1*a1c+a0w+1*a1w>=3.14;
lincon  a0c+2*a1c+a0w+2*a1w>=3.14;
lincon  a0c+1*a1c+a0w+1*a1w>=3.14;
lincon  a0c+2*a1c+a0w+2*a1w>=3.14;
lincon  a0c+11*a1c+a0w+11*a1w>=4.62;
lincon  a0c+2*a1c+a0w+2*a1w>=3.14;
lincon  a0c+1*a1c+a0w+1*a1w>=3.14;
lincon  a0c+5*a1c+a0w+5*a1w>=4.09;
lincon  a0c+9*a1c+a0w+9*a1w>=4.38;
y=a0w*35+260*a1w;
run;
```

```
lincon  a0c+1*a1c+a0w+1*a1w>=3.14;
lincon  a0c+1*a1c+a0w+1*a1w>=3.14;
lincon  a0c+5*a1c+a0w+5*a1w>=4.09;
lincon  a0c+2*a1c+a0w+2*a1w>=3.14;
lincon  a0c+1*a1c+a0w+1*a1w>=3.14;
lincon  a0c+5*a1c+a0w+5*a1w>=4.09;
lincon  a0c+1*a1c+a0w+1*a1w>=3.14;
lincon  a0c+13*a1c+a0w+13*a1w>=4.94;
lincon  a0c+13*a1c+a0w+13*a1w>=4.94;
lincon  a0c+11*a1c+a0w+11*a1w>=4.62;
lincon  a0c+3*a1c+a0w+3*a1w>=3.69;
lincon  a0c+3*a1c+a0w+3*a1w>=3.69;
lincon  a0c+3*a1c+a0w+3*a1w>=3.69;
lincon  a0c+5*a1c+a0w+5*a1w>=4.09;
lincon  a0c+2*a1c+a0w+2*a1w>=3.14;
lincon  a0c+3*a1c+a0w+3*a1w>=3.69;
lincon  a0c+2*a1c+a0w+2*a1w>=3.14;
lincon  a0c+5*a1c+a0w+5*a1w>=4.09;
lincon  a0c+1*a1c+a0w+1*a1w>=3.14;
lincon  a0c+9*a1c+a0w+9*a1w>=4.38;
lincon  a0c+1*a1c+a0w+1*a1w>=3.14;
lincon  a0c+9*a1c+a0w+9*a1w>=4.09;
lincon  a0c+13*a1c+a0w+13*a1w>=4.94;
```

The Algorithm of Exponential Calculation

The algorithm is an SAS programming method for adding data and measuring the quality of predicted and original data using interclass correlation (ICC). Two variables, Read1 and Read2, as well as pid, make up the data.

```
    data test_data;
    input Read1 Read2 pid;
    cards;
      25      23      1
      25      23      2
      44      40      3
      66      60      4
      88      80      5
      111     101     6
      154     140     7
    ;
run;
data test_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;
```

Results

The parameter estimation for Fuzzy Least Square Exponential Growth is shown in the results below (Table 3).

Table 3: Optimization Results

No	Parameter	Estimate	Gradient Objective Function	Active Bound Constraint
1	a0c	3.140000	0	
2	a0w	0.271429	35.000000	
3	a1c	0.135714	0	
4	a1w	0	260.000000	Lower BC

Value of Objective Function = 9.5

Parameter estimates are given: a0c=3.140000, a0w=0.271429, a1c=0.135714 and a1w= 0. For an exponential model, the fuzzy lower limit of prediction is calculated using the following equation:

$$\ln y = (3.140000-0.271429) +(0.135714-0)\ x \quad (4)$$

$\ln y= 2.868571+0.135714x$

and the upper limit of prediction interval for the exponential model is computed using the equation:

$\ln y = (3.140000+0.271429) +(0.135714+0)\ x$

$$\ln y= 3.411429+0.135714x \quad (5)$$

Evaluating Closeness Agreement of Observations using Fuzzy Regression Equation

Table 4 shows the predicted values based on the fuzzy regression equation (5)

Table 4: Original vs Predicted Data

Original Data*	Predicted Data*
23	26
23	39
40	45
60	59
80	102
101	134
140	176

We used the following syntax to calculate regression based on fuzzy regression for exponential growth using ICC:

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

Table 5: Intraclass Correlation Coefficient (ICC)

Reading	Covariate Parameter	Estimate	ICC
1	Intercept	2342.79	
2	Residual	225.75	0.91212

1= Predicted Data
2=Original Data

Table 5 shows that the ICC value is 0.91212, which indicates that predicted and original data are almost perfect agreement.

Evaluating closeness agreement of observations using linear equation

Using ICC, the output displays the quality of the predicted data and the original data. Table 6 illustrates the ICC results in the content of agreement studies based on predicted and original data.

Table 6: Intraclass Correlation Coefficient (ICC)

Reading	Covariate Parameter	Estimate	ICC
1	Intercept	2066.00	
2	Residual	30.0000	0.98569

1= Predicted Data
2=Original Data

The table 6 shows that the value of ICC is 0.98569 and this agreement is almost perfect or almost complete reliability for evaluating the quality of predicted data with the original data.

Discussion

This paper gives an explanation for the agreement between the two methods on microorganism data set in order to determine if they can be used interchangeable and to estimate the fuzzy regression and bootstrap method parameter of the models. Using linear equation, the value of ICC is 0.98569 while the value of ICC using fuzzy regression equation 0.91212. The results of this study show that an ICC of 0.80 or higher indicates a relationship between two variables that is significantly below what is reasonable for classifying it as outstanding, whether using the linear equation or fuzzy regression equation. The diagnostic and predictive situations that researchers and practitioners in podiatry would face in this case. This indicates that there is a good correlation between the values.

According to (Robert Trevethan, 2017), this is supported by the 95 percent confidence interval, which in this case ranges from 0.60 to 0.80, implying that, while ICC from similar samples would differ little from the obtained ICC of 0.80, none of them would be as high as the value of 0.90 that would qualify them as acceptable within a clinical context. Most studies of intrarater and interrater reliability appear to ignore these sources of variation, despite the fact that they virtually always play a role. The ICC of 0.90 in this study is proof of that. Previous study from (Rohim et al., 2020), reported that the exponential regression method by adding bootstrap and fuzzy techniques is more efficient than exponential regression technique.

Limitations and Recommendations

The correlation coefficient is a widely used tool for determining measuring method agreement. It is easy to calculate, but there are some limitations of study. The first restriction is that ICC does not provide any type information and is particularly sensitive to the study's range of values. Therefore, it is highly preferable to use Bland-Altman plots instead, as these reveal both systematics and random errors. In the case of repeated measurements and calibrations, Bland- Altman plots are also preferred.

Conclusion

In conclusion, the overall results of this study show that an ICC of 0.80 and above represents a relationship between two variables that is well below that seems appropriates for describing it as excellent. This article will instructs readers on what to look for when encountering ICC in a piece of writing. This study may also motivate researchers to reveal more about their ICC studies, as well as encourage reviewers and editors to demand complete and precise information from authors.

Declaration of Conflicting Interest

The author(s) declared no potential conflict of interest with respect to the research, authorship and publication of this research.

Funding

The authors received no financial support for the research or publication of this research.

Ethical Consideration

This study was approved by the Human Research Ethics Committee USM (HREC). The number of approval is USM/JEPEM/18080391, approval date is 27th August 2018.

References

1) Streiner DL, Norman GR. Health Measurement Scales (Ch.8). 4th edn. Oxford University Press: Oxford 2008.

2) Koo TK, Li MY. A Guideline of Selecting and Reporting Intraclass Correlation Coefficients for Reliability Research. J Chiropr Med. 2016; 15: 155–163. https://doi.org/10.1016/j.jcm.2016.02.012 PMID: 27330520

3) Franco, A., Malhotra, N., Simonovits, G.: Publication bias in the social sciences: unlocking the file drawer. Science 345, 1502–1505 (2014). doi:10.1126/science.1255484

4) Hallgren, K.A.: Computing inter-rater reliability for observational data: an overview and tutorial. Tutor Quant Methods Psychol 8(1), 23–34 (2012)

5) Hayen, A., Dennis, R., Finch, C.: Determining the intra- and inter-observer reliability of screening tools used in sports injury research. J Sci Med Sport 10, 201–210 (2007). doi:10.1016/j.jsams.2006.09.002

6) Gisev, N., Bell. J., and Chen T. (2013). Integrate agreement and interrater reliability: Key Concept, Approaches, and Applications. Res Soc Admin Pharm. Vol. 9(3) pp:330-338.

7) Kreft, J.U., Booth, G., and Wimpenny, J.W.T. (1998). Racism a Simulator for Individual-based Modeling of Bacterial Colony Growth. Microbiology. Vol.144, pp: 3575-3287.

8) Mclinden, C., Khadka, J., and Pesudovs, K. (2011). Statistical Methods for Conducting Agreement (Comparison of Clinical Tests) and Precision (Repeatability or reproducibility) studies in optometry and ophthalmology. Ophthalmic Physiol Opt. Vol. 31, pp: 330-338.

9) Kim., KJ, Moskowitz H., & Koksalan M. (1996). Fuzzy versus Statistical Linear Regression. European Journal of Operation Research, 92(2), 417-437.

10) Bargiela, A., Pedrycz W., & Nakashima T. (2007). Multiple Regression with Fuzzy Data. Fuzzy Set and and systems. 158 (19), 2169-2188.

11) .Lu N, Chen T, Wu P, Gunzler D, Zhang H, He H, et al. Functonal response models for intraclass correlation coefficients. Applied Statstcs. 2014; 41: 2539-2556. doi: htp://dx.doi.org /10.1080/02664763.2014.920780

12) Koo, T.K, and Li M.Y (2016). A Guideline of Selecting and Reporting Intraclass Correlation Coefficients for Reliability Research. Journal of Chiropractic Medicine. Vol. 15 (2), pp: 155–63.

13) Rohim RAA, Ahmad WMA, Ismail NH, Yaaqob MA, Alam MK and Ghazali FMM (2020). Study of Oral Lactobacillus towards Developing a Comprehensive Structured for Integrated Exponential Regression Model. Bangladesh Journal of Medical Science. Vol 19. No.3, pp: 552-557.

14) Rohim RAA, Ahmad WMA, Ismail NH, Ghazali FMM and Alam MK and (2020). Modeling the Growth of Bacteria Streptococcus Sobrinus using Exponential Regression.Pesqui Bras Odontopediatria Clin Integr.20; e5380.

15) Robert Trevethan (2017). Intraclass Correlation Coefficients: Clearing the Air, Extending Some Cautions, and Making Some Requests. Health Services and Outcomes Research Methodology, 2017, 17, 127–143. http://dx.doi.org/10.1007/s10742-016-0156-6.