Fuzzy regression model with Bayesian approach and its application to public health data

Wan Muhamad Amir Wan Ahmad1, Nurul Asyikin Nizam Akbar2, Farah Muna Mohamad Ghazali1, Nor Farid Mohd Noor1, Nor Azlida Aleng3

1School of Dental Sciences, Health Campus, Universiti Sains Malaysia, 16150 Kubang Kerian, Kota Bharu, Kelantan, Malaysia
2Department of Hematology & Transfusion Medicine Unit, Health Campus, Universiti Sains Malaysia, 16150 Kubang Kerian, Kota Bharu, Kelantan, Malaysia
3Faculty of Ocean Engineering Technology and Informatics, Universiti Malaysia Terengganu, Kuala Terengganu, Terengganu, Malaysia.

*Corresponding E-mail: wmamir@usm.my

ARTICLE INFO

Article History
Received 11 April 2021
Revised 20 January 2022
Accepted 1 May 2022

Keywords
Bootstrap
Bayesian linear regression
Fuzzy regression

How to cite this article:
Wan Ahmad, W. M., Akbar, N. A. N., Ghazali, F. M. M., Mohd Noor, N. F., & Aleng, N. A. (2022). Fuzzy regression model with Bayesian approach and its application to public health data. Bulletin of Applied Mathematics and Mathematics Education, 2(1), 19-32.

The application of the Bayesian Linear Regression (BLR) and Fuzzy Bayesian Linear Regression method through the SAS algorithm is the focus of this paper. As an alternative method of data analysis in biostatistics, this modified method can be used. This modified method includes a bootstrapping technique, residual normality checking and some Bayesian Linear Regression Modeling (BLR) enhancement through Fuzzy Bayesian Linear Regression. We illustrated the application of the algorithm for Bayesian Linear Regression (BLR) and Fuzzy Bayesian Linear Regression in this paper.

This is an open access article under the CC-BY-SA license.

ABSTRACT

ARTICLE

Introduction
Bayesian Linear Regression (BLR) analysis is an approach to linear regression in which the statistical analysis is undertaken within the context of Bayesian inference. This technique can be applied to forecast the value of the response variables (dependent) when given any value of the predictor variables (independent variables). A general regression model is given by \( y_i = E(y_i | x_i) + e_i \), where \( i = 1,2,3,...,n \) denoting an observation of a subject, \( y_i \) is the response variable and \( x_i \) is a \( k \times 1 \) vector of independent variables. \( E(y_i | x_i) \) is the expectation of \( y_i \) conditional on \( x_i \), and \( e_i \) is the error term. This paper provides an algorithm for Bayesian Multiple Linear Regressions (BMLR) in SAS (Diem Ngo & La Puente, 2012).
Assume a BLR model where the response vector \( y \) has dimension \( n \times 1 \) and follows a multivariate Gaussian distribution with mean \( X\beta \) and covariance matrix \( \sigma^2 I \), where \( X \) the design matrix has dimension \( n \times p \), \( \beta \) contains the \( p \) regression coefficients, \( \sigma^2 \) is the common variance of the observational and \( I \) is an \( n \times n \) identity matrix. That is, \( y \sim N(X\beta, \sigma^2) \). In the Bayesian approach, the data are supplemented with additional information in the form of a prior probability distribution. The prior belief about the parameters is combined with the data's likelihood function according to Bayes theorem to yield the posterior belief about the parameters \( \beta \) and \( \sigma \) (Gelman et al., 2013; Gelman & Hill, 2006). Data transformation tools are commonly used to improve the normality of distribution and equalizing variance to meet assumptions and improve effect sizes, thus constituting important aspects of data cleaning and preparing for statistical analyses. The traditional transformations that are commonly discussed include: adding constants, square root, converting to logarithmic scales, inverting and reflecting, and applying trigonometric transformations such as sine wave transformations (Osborne, 2010). The study uses Box-Cox transformation. The form of Box-Cox transformation is as below.

\[
y(\lambda) = \begin{cases} 
y' - 1 & \text{if } \lambda \neq 0 \\
\frac{1}{\lambda} \log y & \text{if } \lambda = 0
\end{cases}
\]

where, \( y \) is the observation data and \( \lambda \) is the model parameter. The optimal value of \( \lambda \) was determined and this study used, \( \lambda = 2 \). The example for the application of the method discussed by using SAS language computer software is provided (Osborne, 2010).

The bootstrap methods begin with original data or sample that is taken from the population, then calculated as sample statistics. The next step is to copy the original sample several times to create a pseudo-population with replacement by using the empirical density function (EDF), (Efron, Bradley and Tibshirani, 1993). The benefit of using bootstrap is its capability to develop a sample the same size of the original, which may include an observation several times while omitting other observations. The bootstrap method draws the samples with replacement and calculates statistics for each sample (it stores these statistics and creates a distribution for further analysis). After finalizing the bootstrap, the data is analyzed for mean, standard deviation, confidence intervals, and any other evidence of replication (Cassel, 2010; Jung, Jhun, & Lee, 2005; Higgins, 2005). In applying the bootstrap method, the original findings from the empirical test were replicated several times to meet the research requirement. As an example, for 1000 observations (original data), the analysis is performed by using a statistical linear model. The analysis results of beta coefficients and r-squared are obtained, followed by the application of the bootstrapping method to the selected data. In applying the bootstrap method, a sample of 23 observations was replicated 6 times (this is equal to 115 observations). The analysis from the statistical linear model, the beta coefficients, and r-squared values of the bootstrap method was compared to the original results. The bootstrap method findings depict the average beta coefficients and r-squared values that are similar to the original findings, from where it was replicated. Interestingly, the bootstrap method provides another noble opportunity for a further comprehensive study of science and non-science discipline. A fuzzy regression model can be written as \( Y = Z_0 + Z_{x_1} + Z_{x_2} + \ldots + Z_{x_p} \), 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 fuzzy, the same which also applies to the parameters. We aim to estimate these parameters. In further discussion, \( Z_i \)'s are assumed as symmetric fuzzy numbers.
which can be presented by intervals. For example, $Z_i$ can be expressed as a fuzzy set given by $Z_i = \langle a_i, a_w \rangle$ where $a_i$ is center and $a_w$ is radius or vagueness associated. The fuzzy set above reflects the confidence in the regression coefficients around $a_i$ in terms of symmetric triangular membership functions. The application of this method should be given more attention when the underlying phenomenon is fuzzy, which indicates that the response variable is fuzzy. $T$, the relationship is also considered to be fuzzy. This $Z_i = \langle a_i, a_w \rangle$ can be written as $Z_i = [a_i, a_w]$ with $a_i = a_w - a_i$ and $a_w = a_w - a_i$. In fuzzy regression methodology, parameters are estimated by minimizing total vagueness in the model. $y_i = Z_i + Z_i x_i + Z_i + \ldots + Z_i x_n$. Using $Z_i = \langle a_i, a_w \rangle$, it can be written $y_i = a_i + a_w x_i + \cdots + a_w x_n$. Thus, this can be written as $y_i = a_i + a_w x_i + \cdots + a_w x_n$. Then parameters $Z_i$ are estimated by minimizing the quantity, which is the total vagueness of the model-data set combination, subject to the constraint that each data point must fall within the estimated value of the response variable. This can be visualized as

the following linear programming problem, minimized $\sum_{j=1}^{n} a_i x_j + a_w x_j$ subject to

$$\left( a_i + \sum_{j=1}^{n} a_w x_j \right) - \left( a_w + \sum_{j=1}^{n} a_w x_j \right) \leq Y_i$$

and

$$\left( a_i + \sum_{j=1}^{n} a_w x_j \right) - \left( a_w + \sum_{j=1}^{n} a_w x_j \right) \geq Y_i$$

and $a_w \geq 0$. The simple procedure is commonly used to solve the linear programming problem. (Kacprzyk & Fedrizzi, 1992). Data for this study is a sample which is composed of four variables.

**Method**

**Sample size determination**

The sample size for multiple regression analysis was calculated by using G*power with effect size = 0.15, 0.05, power of the study = 0.80, and the number of predictors were three. The minimum sample size requires is 77 respondents.

| Analysis: | A priori: Compute required sample size |
|-----------|---------------------------------------|
| **Input:**| Effect size $f^2$ = 0.15 |
|           | $\alpha$ err prob = 0.05 |
|           | Power (1-$\beta$ err prob) = 0.80 |
|           | Number of predictors = 3 |
| **Output:**| Noncentrality parameter $\lambda$ = 11.55000 |
|           | Critical F = 2.730 |
|           | Numerator, df. = 3 |
|           | Denominator, df. = 73 |
|           | Total sample size = 77 |
|           | Actual power = 0.80 |
We used variables as shown in Table 1.

| Num. | Variables | Explanation of user variables |
|------|-----------|-------------------------------|
| 1.   | Choltot   | Total Cholesterol             |
| 2.   | Hdl       | HDL Cholesterol               |
| 3    | Trig      | Triglycerides                 |
| 4    | Waist     | Waist circumferences          |

Algorithm and flow chart for modified Bayesian linear regression analysis method
The algorithm for modified Bayesian linear regression analysis method is presented in Figure 1.

```
/*ADDITION BOOTSTRAPPING ALGORITHM TO THE METHOD */
%MACRO bootstrap(data=_last_, booted=booted, boots=2, seed=1234);
   DATA &booted;
      ** randomly picks an integer from 1 to n;**
      pickobs = INT(RANUNI(&seed)*n)+1;
      ** POINT tells SAS to read value pickobs**
      ** NOBS sets n to number of obs in &Data**;
      ** when the point option is used SAS will loop through the data step**
      forever;
      SET &data POINT = pickobs NOBS = n;
      ** saves number of current bootstrap;**
      REPPLICATE=int(i/n)+1;
      i+1;
```

Figure 1. Modified Bayesian linear regression analysis
** stop will leave data set when n*boots obs have been created;
IF i > n*boots THEN STOP;
RUN;

%MEND bootstrap;

Data Cholesterol;
Input choltot Hdl Trig Waist;
Cards;
181 46 76 98.0
220 39 151 94.0
220 39 151 94.0
213 45 123 95.0
179 42 139 81.0
179 42 139 81.0
114 42 62 104.0
114 42 62 104.0
267 71 122 91.5
267 71 122 91.5
235 73 91 96.5
247 55 85 92.0
199 57 126 116.5
199 57 126 116.5
162 45 100 88.0
237 70 222 91.5
207 66 81 85.0
202 49 118 106.5
184 43 98 90.0
299 56 207 113.0
184 47 118 95.0
181 92 71 97.5
220 39 151 94.0
220 39 151 94.0
180 47 58 97.0
179 42 139 81.0
179 42 139 81.0
179 42 139 81.0
179 42 139 81.0
114 42 62 104.0
114 42 62 104.0
114 42 62 104.0
247 55 85 92.0
199 57 126 116.5
199 57 126 116.5
162 45 100 88.0
210 32 193 95.5
237 70 222 91.5
207 66 81 85.0
202 49 118 106.5
299 56 207 113.0
184 47 118 95.0

; ods rtf file='abc.rtf' style=journal;

**generate bootstrap sample;

%bootstrap(data=Cholesterol, boots=2);
run;
/*PRINT DATA */
proc print data=booted;
run;
ods rtf close;

/* RESIDUAL NORMALITY CHECKING*/
Data Booted;
Input chol tot bayes hdl trig waist;
Cards;
207.54 73.00 91.00 96.50
190.94 45.00 123.00 95.00
193.40 66.00 81.00 85.00
202.99 42.00 139.00 81.00
207.54 73.00 91.00 96.50
190.94 45.00 123.00 95.00
203.47 39.00 151.00 94.00
202.99 42.00 139.00 81.00
170.81 43.00 98.00 90.00
137.86 42.00 62.00 104.00
203.47 39.00 151.00 94.00
264.48 56.00 207.00 113.00
190.94 45.00 123.00 95.00
193.40 66.00 81.00 85.00
264.48 56.00 207.00 113.00
137.86 42.00 62.00 104.00
193.40 66.00 81.00 85.00
303.12 70.00 222.00 116.50
204.00 57.00 126.00 97.50
220.11 92.00 71.00 97.50
178.03 55.00 85.00 92.00
175.88 45.00 100.00 88.00
207.54 73.00 91.00 96.50
170.81 43.00 98.00 90.00
202.99 42.00 139.00 81.00
137.86 42.00 62.00 104.00
156.15 46.00 76.00 98.00
220.11 92.00 71.00 97.50
193.40 66.00 81.00 85.00
229.48 71.00 122.00 91.50
204.00 57.00 126.00 116.50
137.86 42.00 62.00 104.00
202.99 42.00 139.00 81.00
203.47 39.00 151.00 94.00
202.99 42.00 139.00 81.00
137.86 42.00 62.00 104.00
189.41 49.00 118.00 94.00
137.86 42.00 62.00 104.00
175.88 45.00 100.00 88.00
137.86 42.00 62.00 104.00
224.26 32.00 193.00 95.50
202.99 42.00 139.00 81.00
204.00 57.00 126.00 116.50
264.48 56.00 207.00 113.00
144.42 47.00 58.00 97.00
204.00 57.00 126.00 116.50
202.99 42.00 139.00 81.00
203.47 39.00 151.00 94.00
137.86 42.00 62.00 104.00
202.99 42.00 139.00 81.00
207.54 73.00 91.00 96.50
175.88 45.00 100.00 88.00
Fuzzy regression model (Wan Ahmad, Akbar, Ghazali, Mohd Noor, Awang, Aleng)
| 193.40 | 66.00 | 81.00 | 85.00 |
| 264.48 | 56.00 | 207.00 | 113.00 |
| 137.86 | 42.00 | 62.00 | 104.00 |
| 193.40 | 66.00 | 81.00 | 85.00 |
| 303.12 | 70.00 | 222.00 | 91.50 |
| 204.00 | 57.00 | 126.00 | 116.50 |
| 220.11 | 92.00 | 71.00 | 97.50 |
| 178.03 | 55.00 | 85.00 | 92.00 |
| 175.88 | 45.00 | 100.00 | 88.00 |
| 207.54 | 73.00 | 91.00 | 96.50 |
| 170.81 | 43.00 | 98.00 | 90.00 |
| 202.99 | 42.00 | 139.00 | 81.00 |
| 137.86 | 42.00 | 62.00 | 104.00 |
| 156.15 | 46.00 | 76.00 | 98.00 |
| 220.11 | 92.00 | 71.00 | 97.50 |
| 193.40 | 66.00 | 81.00 | 85.00 |
| 229.48 | 71.00 | 122.00 | 91.50 |
| 204.00 | 57.00 | 126.00 | 116.50 |
| 137.86 | 42.00 | 62.00 | 104.00 |
| 202.99 | 42.00 | 139.00 | 81.00 |
| 203.47 | 39.00 | 151.00 | 94.00 |
| 189.41 | 49.00 | 118.00 | 106.50 |
| 178.03 | 55.00 | 85.00 | 92.00 |
| 202.99 | 42.00 | 139.00 | 81.00 |
| 190.12 | 47.00 | 118.00 | 95.00 |
| 144.42 | 47.00 | 58.00 | 97.00 |
| 204.00 | 57.00 | 126.00 | 116.50 |
| 202.99 | 42.00 | 139.00 | 81.00 |
| 203.47 | 39.00 | 151.00 | 94.00 |
| 137.86 | 42.00 | 62.00 | 104.00 |
| 175.88 | 45.00 | 100.00 | 88.00 |
| 202.99 | 42.00 | 139.00 | 81.00 |
| 224.26 | 32.00 | 193.00 | 95.50 |
| 137.86 | 42.00 | 62.00 | 104.00 |
| 190.12 | 47.00 | 118.00 | 95.00 |
| 189.41 | 49.00 | 118.00 | 106.50 |
| 178.03 | 55.00 | 85.00 | 92.00 |
| 203.47 | 39.00 | 151.00 | 94.00 |
| 137.86 | 42.00 | 62.00 | 104.00 |
| 175.88 | 45.00 | 100.00 | 88.00 |
| 202.99 | 42.00 | 139.00 | 81.00 |
| 204.00 | 57.00 | 126.00 | 116.50 |
| 264.48 | 56.00 | 207.00 | 113.00 |
| 144.42 | 47.00 | 58.00 | 97.00 |
| 204.00 | 57.00 | 126.00 | 116.50 |
| 202.99 | 42.00 | 139.00 | 81.00 |
| 203.47 | 39.00 | 151.00 | 94.00 |
| 137.86 | 42.00 | 62.00 | 104.00 |
| 175.88 | 45.00 | 100.00 | 88.00 |
| 178.03 | 55.00 | 85.00 | 92.00 |
| 137.86 | 42.00 | 62.00 | 104.00 |
| 193.40 | 66.00 | 81.00 | 85.00 |
| 207.54 | 73.00 | 91.00 | 96.50 |
| 175.88 | 45.00 | 100.00 | 88.00 |
| 178.03 | 55.00 | 85.00 | 92.00 |
| 137.86 | 42.00 | 62.00 | 104.00 |
| 170.81 | 43.00 | 98.00 | 90.00 |
| 137.86 | 42.00 | 62.00 | 104.00 |
| 204.00 | 57.00 | 126.00 | 116.50 |
| 170.81 | 43.00 | 98.00 | 90.00 |
| 137.86 | 42.00 | 62.00 | 104.00 |
| 204.00 | 57.00 | 126.00 | 116.50 |
```
204.00 57.00  126.00 116.50
303.12 70.00  222.00  91.50
203.47 39.00  151.00  94.00
190.12 47.00  118.00  95.00

; run;
ods rtf file='abc.rtf' style=journal;
ods graphics on;
proc reg data=Booted plots=all;
run;
proc genmod data=Booted;
model choltotbayesian = hdl trig waist / dist=normal link=identity;
bayes seed=1 OutPost=Post diagnostics=all summary=all;
run;
ods graphics off;
ods rtf close;
run;

/* BAYESIAN FUZZY REGRESSION*/

Title 'Linear programming';
data plant;
  input choltotbayesian hdl trig waist;
datalines;
  207.54 73.00  91.00  96.50
  190.94 45.00  123.00  95.00
  193.40 66.00  81.00  85.00
  202.99 42.00  139.00  81.00
  207.54 73.00  91.00  96.50
  190.94 45.00  123.00  95.00
  203.47 39.00  151.00  94.00
  202.99 42.00  139.00  81.00
  170.81 43.00  98.00  90.00
  137.86 42.00  62.00  104.00
  203.47 39.00  151.00  94.00
  264.48 56.00  207.00  113.00
  190.94 45.00  123.00  95.00
  193.40 66.00  81.00  85.00
  264.48 56.00  207.00  113.00
  137.86 42.00  62.00  104.00
  193.40 66.00  81.00  85.00
  303.12 70.00  222.00  91.50
  204.00 57.00  126.00 116.50
  220.11 92.00  71.00  97.50
  178.03 55.00  85.00  92.00
  175.86 45.00 100.00  88.00
  207.54 73.00  91.00  96.50
  170.81 43.00  98.00  90.00
  202.99 42.00  139.00  81.00
  137.86 42.00  62.00  104.00
  156.15 46.00  76.00  98.00
  220.11 92.00  71.00  97.50
  193.40 66.00  81.00  85.00
  229.48 71.00  122.00  91.50
  204.00 57.00  126.00 116.50
  137.86 42.00  62.00  104.00
  202.99 42.00  139.00  81.00
  203.47 39.00  151.00  94.00
```
proc optmodel;
    set j = 1..84;
    number choltotbayesian{j}, hdl{j}, trig{j} waist{j};
    read data plant into [_n_]choltotbayesian hdl trig waist;

    /*Print choltotbayesian  hdl  trig waist*/
    print choltotbayesian  hdl  trig waist*/
    print choltotbayesian  hdl  trig waist;
    number n init 8;
    /*Total number of Observations*/
    /*Decision Variables*/
Results and Discussion

Results from Bayesian multiple linear regression

The results from Bayesian multiple linear regression is presented in Table 2.

| Parameter | Estimate | Standard Error | Wald 95% Confidence Limits |
|-----------|----------|----------------|-----------------------------|
| Intercept | 62.5793  | 0.0028         | 62.5739 - 62.5847           |
| Hdl ($X_1$) | 1.4686   | 0.0000         | 1.4685 - 1.4686             |
| Trig ($X_2$) | 0.7511   | 0.0000         | 0.7511 - 0.7511             |
| Waist ($X_3$) | -0.3170  | 0.0000         | -0.3170 - -0.3169           |

with Choltotbayesian($Y$)

Multiple Bayesian Linear Regression (MBLR) is given as follows:

$$ (Y) = 62.5793 + 1.4686 (X_1) + 0.7511 (X_2) -0.3170 (X_3) $$

with

($X_1$) is High Density Lipoprotein reading
($X_2$) is a Triglycerides reading
($X_3$) is Waist reading

Fitted Bayesian Multiple linear Regression with standard error is given as follows:

$$ (Y) = 62.5793 + 1.4686 (X_1) + 0.7511 (X_2) $$

Std. Error (0.0028) (0.0000) (0.0000) 
-0.3170 ($X_3$)............................ (2.1)
Std. error (0.0000)
Upper or lower limits of prediction interval are computed from the prediction equation (2.1) by taking the coefficient as their corresponding estimated values plus or minus standard error (See Table 3).

**Upper limits**

\[
(\hat{y}) = 62.5821 + 1.4686 \cdot (x_1) + 0.7511 \cdot (x_2) - 0.3170 \cdot (x_3) \]

\[(2.2)\]

**Lower limits**

\[
(\hat{y}) = 62.5765 + 1.4686 \cdot (x_1) + 0.7511 \cdot (x_2) - 0.3170 \cdot (x_3) \]

\[(2.3)\]

| Table 3. Lower and upper limit with their width for Bayesian multiple linear regression |
|---------------------------------|---------------------------------|----------------|
| Lower limit | Upper limit | Width |
| 207.544 | 207.550 | .0056 |
| 190.934 | 190.939 | .0056 |
| 193.398 | 193.404 | .0056 |
| 202.984 | 202.989 | .0056 |
| 207.544 | 207.550 | .0056 |
| 190.934 | 190.939 | .0056 |
| 203.470 | 203.476 | .0056 |
| \| | \| | \|
| 303.117 | 303.123 | .0056 |
| 203.470 | 203.476 | .0056 |
| 190.116 | 190.121 | .0056 |
| Average width | 0.0056 |

Results from fitted model for Fuzzy regression is presented in Table 4.

**Table 4. Value of center (AC) and radius (AW)**

| [1] | AC  | AW    |
|-----|-----|-------|
| 1   | 62.65633 | 0.0015848 |
| 2   | 1.46796  | 0.0000000 |
| 3   | 0.75085  | 0.0000000 |
| 4   | -0.31716 | 0.0000000 |

Fitted model for fuzzy regression (FR) for Cholotobaysian \((\hat{y}) =<62.6547452, 0.0015848>+<1.46796, 0.0000000> \cdot \text{Hdl} +<0.75085, 0.0000000> \cdot \text{Trig} +<-0.31716, 0.0000000> \cdot \text{Waist} \]

\[(2.4)\]

Upper or lower limits of prediction intervals are computed from the prediction equation (2.4) by taking the coefficient as their corresponding estimated values plus or minus standard error (See Table 5).

**Upper limits**

\[
Y = <62.6547452, 0.0015848>+<1.46796, 0.0000000> \cdot (X_1) +<0.75085, 0.0000000> \cdot (X_2) +<-0.31716, 0.0000000> \cdot (X_3) \]

\[(2.5)\]

**Lower limits**

\[
Y = <62.6579148, 0.0015848>+<1.46796, 0.0000000> \cdot (X_1) +<0.75085, 0.0000000> \cdot (X_2) +<-0.31716, 0.0000000> \cdot (X_3) \]

\[(2.6)\]
Table 5. Lower and Upper Limit with Their Width for Fitted Model for Fuzzy Regression

| Lower limit | Upper limit | Width  |
|-------------|-------------|--------|
| 202.990     | 202.987     | 0.0032 |
| 207.540     | 207.537     | 0.0032 |
| 190.940     | 190.937     | 0.0032 |
| 203.474     | 203.470     | 0.0032 |
| 202.990     | 202.987     | 0.0032 |
| 170.819     | 170.816     | 0.0032 |
| 137.880     | 137.877     | 0.0032 |
| 303.084     | 303.081     | 0.0032 |
| 203.474     | 203.470     | 0.0032 |
| 190.122     | 190.119     | 0.0032 |

Average width 0.003170

The width of prediction intervals concerning the Bayesian multiple linear regression model and the Bayesian fuzzy regression model corresponding to each set of observed explanatory variables is computed in SPSS and the results are reported in Table 5. From this table, the average width for the former was found to be 0.005600, while that of the latter was only 0.003170, thereby indicating the superiority of fuzzy regression methodology.

Conclusion

This paper presents an algorithm and illustrated the procedure of modeling by using modified Bayesian linear regression through SAS language. Our aim is to share the algorithm and also provide the researcher with an alternative programming that suitable for a small sample size. This proposed method can be applied to small sample size data, especially when limited data is obtained, for example in public health.

Acknowledgement

The authors would like to express their gratitude to Universiti Sains Malaysia (USM) for providing International Book Grant (Grant No.304/PPSG/6150179/T150), School of Dental Sciences.

References

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 333-2012, 1-12.

Cassel, D. L. (2010). Bootstrap Mania: Re sampling the SAS. SAS Global Forum 2010: Statistics and Data Analysis. Paper 268-279. Proceedings of the SAS Global Forum 2010 Conference. Cary (NC): SAS Institute Inc.

Jung, B. C., Jhun, M., & Lee, J. W. (2005). Bootstrap Tests for Overdispersion in a Zero-Inflated Poisson Regression Model. Biometrics, 61, 626-629.

Kacprzyk J., & Fedrizzi M. (1992) Fuzzy Regression Analysis, Warsaw: Omnitech Press.

Bradley, E., & Tibshirani, R. J. (1993). An Introduction to the Bootstrap. New York, NY: Chapman and Hall.

Higgins, G. E. (2005). Statistical Significance Testing: The Bootstrapping Method and an Application to Self-Control Theory. The Southwest Journal of Criminal Justice, 2(1), 54-76.

Gelman, A., Carlin, J. B., Stern, H. S., Dunson, D. B., Vehtari, A., & Rubin, D. B. (2013). Bayesian Data
Gelman, A., & Hill, J. (2006). *Data Analysis Using Regression and Multilevel/Hierarchical Models*. Cambridge: Cambridge University Press.

Osborne, J.W. (2010). Improving your data transformations?: Applying the Box-Cox transformation. *Practical Assessment, Research & Evaluation, 15*(12), 1-9.