MODELING EARLY CHILHOOD CARIES (ECC) USING BAYESIAN REGRESSION ANALYSIS APPROACH TO PUBLIC HEALTH DATA


** School of Dental Sciences, Health Campus, Universiti Sains Malaysia (USM), Kubang Kerian, Kelantan, Malaysia
*** School of Informatics and Applied Mathematics, Universiti Malaysia Terengganu (UMT), Kuala Terengganu, Terengganu

*School of Mathematics Sciences, Universiti Sains Malaysia (USM), Minden, Pulau Pinang, Malaysia

Abstract:
This paper focuses on the two main parts, the first part, which was focused on comprehensive methodology building and the second part is an application of Bayesian Linear Regression (BLR) through SAS algorithm using Early Childhood Caries (ECC) datasets. This method totally emphasize on specific modified algorithm which suit the alternative method. Through the gained results, it is proof that this method can be applied to biostatistics field, dental public health and others. In this paper, we discussed and illustrated the methodology algorithm building in details with an application of BLR toward ECC dataset

Key Words: Early Childhood Caries (ECC), Public Health, Bayesian Linear Regression (BLR) & SAS Algorithm.

1. Early Childhood Caries (ECC):
Early childhood caries (ECC) is one of the most common chronic conditions among young children as the teeth erupt and affects the oral health of 71 months of age or younger children today. According to Dye et al., in 2007 ECC remain number one chronic disease in United State. ECC affect the quality of life of children due to dental pain, missing, decayed and filling tooth surfaces in any primary tooth in a preschool-age child (Abu, 2013; Berg & Slayton, 2009). According to the study of National Institute of Dental and Craniofacial Research (2000) in US stated that dental carries in children is five times more likely compared to asthma and seven times more likely compared have hay fever. There are several type of dental carries pattern such as “bottle carries,” “nursing carries” and “baby bottle tooth decay”. Most of the reported cases of early childhood carries is caused by inappropriate bottle feeding. Centers for Disease Control and Prevention conducted a study in 2005 and found that approximately 40% of children having dental carries at 5 years old and 8% at 2 years old. It also revealed by the previous researcher that the dental carries has changed the children quality of life from being ashamed to smile and speak, and also difficult in eating and malnutrition (Barbosa et al., 2008; Zhou et al., 2011; Bener et al., 2013; Martins-Junior et al., 2013; Casamassimo et al, 2009). According Ruhaya et al. in 2012 chronic dental pain may cause avoidance of foods and lead to malnutrition among the children. Lacking of nutrition also can lead to dental carries. But in Malaysia, studies on nutritional status in relation to dental carries are very limited she emphasize that, there were no previous studies exploring the relationships between nutritional status and ECC among preschool children. Most of children are at risk of developing ECC, 1-12% especially in developing country and 70% prevalence of ECC in some developing countries (AAPHD, 2004). According to the study done Hilgers et al., in 2006, the finding shows that carries among children has strong association between obese with high carries and high BMI was associated with high consumption of food and beverages as such candies, chips and many more which lead to increase carries risk among children. These significant finding also was supported by Willershausen et al., in 2006.

2. Sample Size Determination:
Sample size for multiple regression analysis were calculated by using G*power with effect size = 0.02, \( \alpha = 0.05 \), power of the study = 0.68 and number of predictor were 2. The minimum sample size requires is 372 respondents.

Table1: Description of Cholesterol Data

<table>
<thead>
<tr>
<th>No</th>
<th>Variables</th>
<th>Explanation of user variables</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>filling (y)</td>
<td>Number of Filling</td>
</tr>
<tr>
<td>2</td>
<td>sex (x1)</td>
<td>Gender of student</td>
</tr>
<tr>
<td>3</td>
<td>bmi (x2)</td>
<td>Body Mass Index</td>
</tr>
</tbody>
</table>
3. Algorithm and Flow Chart for Modified Bayesian Linear Regression Analysis Method:

Algorithm for Bayesian Regression Model:

Data Cholesterol:
/*BAYESIAN REGRESSION MODEL*/
Data Childhoodcaries;
Input sex bmi caries m filling age;
Cards;
1 14.5 2 2 0 7.00
1 12.7 4 2 0 6.00
2 12.3 4 2 0 7.00
1 13.9 6 2 0 6.00
2 15.6 8 2 0 6.00
2 12.0 2 1 0 6.00
1 15.9 4 1 0 6.00
1 16.6 5 1 0 7.00
1 12.5 6 1 2 7.00
2 14.3 9 1 0 7.00
1 12.4 10 1 0 6.00
2 13.1 10 1 0 7.00
1 14.5 11 1 0 6.00
2 14.4 14 1 0 7.00
2 14.9 14 1 0 7.00
: : : : : :
2 14.6 14 0 0 7.00
2 12.7 14 0 0 6.00
2 14.3 14 0 0 7.00
2 13.6 14 0 0 6.00
2 13.7 14 0 0 7.00
2 13.6 14 0 0 6.00
2 12.3 14 0 0 7.00
2 15.1 14 0 0 6.00
2 12.6 14 0 0 6.00
1 13.9 15 0 0 6.00
### Calculating the P Value of Intercept

The SAS algorithm used to calculate the P value of the intercept is as follows:

\[
\text{Test Statistic} = \frac{-0.6295}{0.2503} = -2.418
\]
Data Childhoodcaries;
/ts = the test statistic, 
df = degrees of freedom;/
ts = -2.418; df = 376;
/*nondirectional P-value */
Pvalue = (1-PROBT(abs(ts),df))*2;
proc print data= Childhoodcaries; run; quit;
run;

Output:

<table>
<thead>
<tr>
<th>Obs</th>
<th>ts</th>
<th>df</th>
<th>P value</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>-2.418</td>
<td>376</td>
<td>0.01608</td>
</tr>
</tbody>
</table>

(ii). Calculation of P Value of Sex Variable:

Test Statistics = \frac{-0.0078 - 0}{0.0334} = -0.2335

Data Childhoodcaries;
/ts = the test statistic, 
df = degrees of freedom;/
ts = -0.2335; df = 376;
/*nondirectional P-value */
Pvalue = (1-PROBT(abs(ts),df))*2;
proc print data= Childhoodcaries; run; quit;
run;

Output:

<table>
<thead>
<tr>
<th>Obs</th>
<th>ts</th>
<th>df</th>
<th>P value</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>-0.2335</td>
<td>376</td>
<td>0.81550</td>
</tr>
</tbody>
</table>

(iii). Calculation of P Value of BMI:

Test Statistics = \frac{0.0211 - 0}{0.0089} = 2.370

Data Childhoodcaries;
/ts = the test statistic, 
df = degrees of freedom;/
ts = 2.370; df = 376;
/*nondirectional P-value */
Pvalue = (1-PROBT(abs(ts),df))*2;
proc print data= Childhoodcaries; run; quit;
run;

Output:

<table>
<thead>
<tr>
<th>Obs</th>
<th>ts</th>
<th>df</th>
<th>P value</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2.370</td>
<td>376</td>
<td>0.01829</td>
</tr>
</tbody>
</table>

(iv). Calculation of P Value of M Variable:

Test Statistics = \frac{0.0011 - 0}{0.0584} = 0.01884

Data Childhoodcaries;
/ts = the test statistic, df = degrees of freedom;/
ts = 0.01884; df = 376;
/*nondirectional P-value */
Pvalue = (1-PROBT(abs(ts),df))*2;
proc print data= Childhoodcaries; run; quit;
run;

Output:

<table>
<thead>
<tr>
<th>Obs</th>
<th>ts</th>
<th>df</th>
<th>P value</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.01884</td>
<td>376</td>
<td>0.98498</td>
</tr>
</tbody>
</table>

(v). Calculation of P Value of Caries Variable:

Test Statistics = \frac{-0.0055 - 0}{0.0033} = -1.6666

Data Childhoodcaries;
/*ts = the test statistic, df = degrees of freedom;/
ts = -1.6666; df = 376;
/*nondirectional P-value */
Pvalue = (1 - PROBT(abs(ts),df))*2;
proc print data= Childhoodcaries; run; quit;
run;

Output:

<table>
<thead>
<tr>
<th>Obs</th>
<th>ts</th>
<th>df</th>
<th>P value</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>-1.6666</td>
<td>376</td>
<td>0.0964</td>
</tr>
</tbody>
</table>

(vi). Calculation of P Value of Age Variable:

Data Childhoodcaries;
/*ts = the test statistic, df = degrees of freedom;*/
ts = 2.1374; df = 376;
/*nondirectional P-value */
Pvalue = (1 - PROBT(abs(ts),df))*2;
proc print data= Childhoodcaries; run; quit; run;

Output:

<table>
<thead>
<tr>
<th>Obs</th>
<th>ts</th>
<th>df</th>
<th>P value</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2.1374</td>
<td>376</td>
<td>0.0332</td>
</tr>
</tbody>
</table>

Data Cholesterol;
/*Response Surface Methodology*/
Data Childhoodcaries;
Input sex bmi caries m filling age;
Cards;
1 14.5  2  2  0  7.00
1 12.7  4  2  0  6.00
2 12.3  4  2  0  7.00
1 13.9  6  2  0  6.00
2 15.6  8  2  0  6.00
2 12.0  2  1  0  6.00
1 15.9  4  1  0  6.00
1 16.6  5  1  0  7.00
1 12.5  6  1  2  7.00
2 14.3  9  1  0  7.00
1 12.4  10 1  0  6.00
2 13.1  10 1  0  7.00
1 14.5  11 1  0  6.00
2 14.4  14 1  0  7.00
2 14.9  14 1  0  7.00
2 14.6  14 0  0  7.00
2 12.7  14 0  0  6.00
2 14.3  14 0  0  7.00
2 14.4  15 0  0  7.00
2 14.1  15 0  0  6.00
1 14.2  16 0  0  6.00
1 13.0  16 0  0  6.00
2 10.8  16 0  0  6.00
2 14.4  16 0  0  7.00
1 16.1  17 0  0  6.00
1 13.1  17 0  0  6.00
1 14.1  17 0  0  6.00
1 15.0  17 0  0  6.00
2 14.0  17 0  0  6.00
2 14.3  17 0  0  6.00
2 13.4  17 0  0  7.00
2 13.3  17 0  0  6.00
1 17.3  18 0  0  7.00
1 14.5  18 0  0  6.00
1 13.4  18 0  0  7.00
4. Results:

Part I: Results from Bayesian Multiple Linear Regression

Table 2: Analysis of Maximum Likelihood Parameter Estimates “Bayesian Multiple linear Regression”

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Estimate</th>
<th>Standard Error</th>
<th>Wald 95% Confidence Limits</th>
<th>P value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>-0.6295</td>
<td>0.2603</td>
<td>-1.1396 -0.1193</td>
<td>0.81550</td>
</tr>
<tr>
<td>Sex</td>
<td>-0.0078</td>
<td>0.0334</td>
<td>-0.0732 0.0577</td>
<td>0.01829*</td>
</tr>
<tr>
<td>Bmi</td>
<td>0.0211</td>
<td>0.0089</td>
<td>0.0036 0.0386</td>
<td></td>
</tr>
<tr>
<td>M</td>
<td>0.0011</td>
<td>0.0584</td>
<td>-0.1134 0.1157</td>
<td></td>
</tr>
<tr>
<td>caries</td>
<td>-0.0055</td>
<td>0.0033</td>
<td>-0.0120 0.0010</td>
<td>0.096400</td>
</tr>
<tr>
<td>Age</td>
<td>0.0731</td>
<td>0.0342</td>
<td>0.0061 0.1401</td>
<td>0.03321*</td>
</tr>
</tbody>
</table>

*Significant at $\alpha = 0.05$. 

Figure 1: Trace Plots for Intercept (a), Sex Variable (b) and BMI Variable (c)
Multiple Bayesian Linear Regression (MBLR) is given as follows:

\[
(Y) = -0.6295 + 0.0078 (x_1) + 0.0211 (x_2) + 0.0011 (x_3) + 0.0055 (x_4) + 0.0731 (x_5)
\]

Where

\(x_1\) is sex
\(x_2\) is body mass index
\(x_3\) is number of missing tooth
\(x_4\) is number of caries
\(x_5\) is number of age

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

\[
(Y) = -0.6295 - 0.0078 (x_1) + 0.0211 (x_2) + 0.0011 (x_3) + 0.0055 (x_4) + 0.0731 (x_5)
\]

Standard Error:

\[
\text{Std. Err.} \begin{pmatrix} 0.6295 \ 0.0078 \ 0.0211 \ 0.0011 \ 0.0055 \ 0.0731 \end{pmatrix}
\]

\[
\text{Std. Err.} \begin{pmatrix} 0.0584 \ 0.0033 \ 0.0342 \end{pmatrix}
\]

Summary and Discussion:

The aim of this paper is to focus the the Bayesian algorithm methodologies building and the factor of ECC caries among children based on the risk factor. We presents an applied regression modelling using Bayesian linear regression approach through SAS language. Table 2 summarize the output for Bayesian
regression analysis for early childhood carries, using modified Bayesian linear regression through SAS language. From the results output, there are two significant factor were revealed as a main factor that contributing to filling among children, in Bachok Kelantan. These two predictor variable were BMI ($\beta_1 = 0.0211, CI: 0.0036, 0.0386$) and age ($\beta_2 = 0.0731, CI: 0.0061, 0.1401$) were statistically significant at level of $\alpha = 0.05$. According to the South Carolina Department of Health and Environmental Control (SC DHEC), 52% of children under the age of eight in South Carolina have experienced tooth decay (2007). The prevalence of ECC among children aged 1-3.5 years in the present study was 69.6% in the tested sample (Abu, 2013).

In this study we found that three other factor which are Missing ($\beta_3 = 0.0011, CI:-0.1134, 0.1157$), Sex ($\beta_4 = -0.0078, CI:-0.0732, 0.0577$) and carries ($\beta_5 = 0.0731, CI:0.0061, 0.1401$) were not statistically significant. This finding was supported by Ruhaya et al. in 2012, there was no significant different between gender and carries experience ($t = 0.422, p > 0.05$). The second aim is to test the alternative algorithm that we build and also provide the researcher with an alternative programming for data analysis purpose especially for public health study.

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

References: