An Alternative Method for Logistic Regression Modeling in Aquatic Benthos: Comparison of Three Different Phases of Multiple Logistic Regressions

Wan Muhammad Luqman Luqman Bin Wan Rosdi, Wan Muhamad Amir W Ahmad, Ruhaya Hasan, Nor Azlida Aleng, Nur Fadhlina Halim, Syerrina Zakaria, Kasypi Mokhtar and Zalila Ali

School of Dental Sciences, Universiti Sains Malaysia (USM), 16150 Kubang Kerian, Kelantan, Malaysia
School of Informatics and Applied Mathematics, Universiti Malaysia Terengganu (UMT), Terengganu Malaysia
School of Maritime Business and Management, Universiti Malaysia Terengganu (UMT), Terengganu Malaysia
School of Mathematics Sciences, Universiti Sains Malaysia (USM), 11800 Minden, Pulau Pinang, Malaysia

Abstract: This paper supplied a complete method of an alternative weighted logistic regression as a technique for analysis through SAS algorithm. This paper also evaluates the marine aquaculture factors that influencing the population of benthos in Bidong Island, Kuala Terengganu, Terengganu, Malaysia. This alternative method is a manipulation technique (using bootstrap) for the small data set and gives the researcher an option to launch the analysis even there is not enough data set. This is an extension of an improvement of the current method by adding weighted and bootstrap to step of building logistic regression model.

Key words: Bootstrap • Aquaculture • Benthos • Weighted and robust regression

INTRODUCTION

Aquaculture includes the farming of finfish or shellfish in a natural habitat with no supplementary food added and with minimum impact on the environment. Nowadays, the intensive farming of marine finfish usually practised in cages or pond, involves the supply of high quality artificial feeds and medication with consequent impacts on the environment. It is mainly due to the release of organic and inorganic nutrients and the release of chemical used in medication. These impacts tend to be most severe in areas with poor water exchange [2]. The objective of this study also is to discuss the improvement between the advanced multiple logistic regression with the original version of multiple logistic regression [3]. We extend the advanced of multiple logistic regressions with combining bootstrap and response surface methodology. Logistic regression is a type of predictive model that can be used when the target variable is a categorical variable with two categories for instance live or die, has cancer or no cancer, having coronary heart disease or not having coronary heart disease, patient survives of dies and many more [4]. The parametric bootstrap method is
recommended for sample size between 50 and 100 for a reliable performance [5, 6]. A recent approach to analyse data with missing values in the covariates is weighted estimating equations and this technique appear to be highly efficient [7].

**Data and Methods:** Data of this study is a sample which composed of six variables. Namely variables are as in Table 1.

Table 1: Description of data.
<table>
<thead>
<tr>
<th>Num.</th>
<th>Code</th>
<th>Explanation of user variables</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.</td>
<td>Y</td>
<td>Feeding Habit</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0 = Suspension Feeder</td>
</tr>
<tr>
<td></td>
<td></td>
<td>1 = Deposit Feeder</td>
</tr>
<tr>
<td>2.</td>
<td>X1</td>
<td>Size</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0 = Small</td>
</tr>
<tr>
<td></td>
<td></td>
<td>1 = Large</td>
</tr>
<tr>
<td>3.</td>
<td>X2</td>
<td>Mobility</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0 = Move</td>
</tr>
<tr>
<td></td>
<td></td>
<td>1 = Don’t move</td>
</tr>
<tr>
<td>4.</td>
<td>X3</td>
<td>Flattened Body</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0 = Flatten</td>
</tr>
<tr>
<td></td>
<td></td>
<td>1 = Not Flatten</td>
</tr>
<tr>
<td>5</td>
<td>X4</td>
<td>Body form</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0 = Rounded</td>
</tr>
<tr>
<td></td>
<td></td>
<td>1 = Not rounded</td>
</tr>
<tr>
<td>6</td>
<td>X5</td>
<td>Life Habitat</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0 = Individual</td>
</tr>
<tr>
<td></td>
<td></td>
<td>1 = Commensalism</td>
</tr>
<tr>
<td>7</td>
<td>X6</td>
<td>Weight</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0 = Less than 0.01g</td>
</tr>
<tr>
<td></td>
<td></td>
<td>1 = More than 0.01g</td>
</tr>
<tr>
<td>8</td>
<td>X7</td>
<td>Distribution</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0 = More than 10 ind.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>1 = Less than 10 ind.</td>
</tr>
</tbody>
</table>

Multiple weighted logistic regression technique was used in the analysis of relationship between variables. The algorithm is given as follows:

**Case I:**

Fig. 1 showed the flow chart of an alternative method of logistic regression procedure.

**CASE I:**

Logistic Regression Programming With SAS

Data Sampling;

input size mobility fbody bform lhabit fhabit weight dist;

data lines;

ods rtf file='robdunc0.rtf' style=journal;

/* Multiple Logistic Regression */

proc logistic data=Sampling;

model fhabit (event='1')= size mobility fbody bform lhabit weight dist;

run;

do rtf close;

Case II

Fig. 1 showed the flow chart of an alternative method of weighted logistic regression procedure.

Fig. 1: Flow chart of an alternative analysis.
CASE II:

Weighted Logistic Regression Programming With SAS

**Data** Sampling;
input size mobility fbody bform lhabit fhabit weight dist;
datelines;
0 1 1 1 1 0 0 0
0 0 0 1 0 0 0 0
0 0 1 1 0 0 0 0
0 0 1 1 0 0 1 0
1 0 1 0 0 0 0 0
0 0 1 0 0 0 0 1
0 0 0 1 0 0 1 0
0 0 0 1 0 0 1 0
;
/* Run The Original Logistic Regression To Get The Residuals*/
proc genmod data=Sampling descending;
class fhabit;
model fhabit = size mobility fbody bform lhabit weight dist /
dist=binomial link=logit;
output out=work.pred reschi=residual;
run;

/* Compute The Absolute And Squared Residuals*/
data work.resid;
set work.pred;
absresid=abs(residual);
sqresid=residual**2;
proc genmod data=work.resid;

/*Run A Regression With The Absolute Residuals Vs. Independent Variables To Get The Estimated Standard Deviation*/
model absresid = size mobility fbody bform lhabit weight dist /
dist=binomial link=logit;
output out=work.s_weights p=s_hat;
run;

/*Compute the weights using the estimated standard deviations*/
data work.s_weights;
set work.s_weights;
s_weight=1/(s_hat**2);
label s_weight = "weights using absolute residuals";

/*Do The Weighted Least Squares Using The Weights From The Estimated Standard Deviation*/

ods rtf close;

Case III

CASE III: Alternative Logistic Regression Programming with combining Weighted and Bootstrap Method

Fig. 1 showed the flow chart of an alternative method of logistic regression with combining weighted and bootstrap method.

**Data** work.s_weights;
set work.s_weights;
s_weight=1/(s_hat**2);
label s_weight = "weights using absolute residuals";

/*Do The Weighted Least Squares Using The Weights From The Estimated Standard Deviation*/
proc genmod data=work.s_weights;
weight s_weight;
model fhabit = size mobility fbody bform lhabit weight dist /
dist=binomial link=logit;
run;
ods rtf close;

CASE III: Alternative Logistic Regression Programming with combining Weighted and Bootstrap Method
Data Sampling;
input size mobility fbody bform lhabit fhabit weight dist;
datalines;
0 1 1 1 1 0 0 0
0 0 0 1 0 0 0 0
0 0 1 1 0 0 0 0
0 0 1 1 0 0 1 0
1 0 1 0 0 0 0 0
0 0 1 0 0 0 0 1
0 0 1 0 0 0 1 0
0 0 0 1 0 0 1 0
% Compute The Absolute And Squared Residuals*/
data work.resid;
set work.pred;
absresid=abs(residual);
sqresid=residual**2;
proc genmod data=work.resid;
/*Run A Regression With The Absolute Residuals Vs. Independent Variables To Get The Estimated Standard Deviation*/
model absresid = size mobility fbody bform lhabit weight dist;
output out=work.s_weights p=s_hat;
run;
/*Compute the weights using the estimated standard deviations*/
data work.s_weights;
set work.s_weights;
s_weight=1/(s_hat**2);
label s_weight = "weights using absolute residuals";
/*Do The Weighted Least Squares Using The Weights From The Estimated Standard Deviation*/
proc genmod data=work.s_weights;
weight s_weight;
model fhabit = size mobility fbody bform lhabit weight dist / dist=binomial link=logit;
run;
ods rtf close;

CONCLUSION

This paper explained on how an alternative programming method of bootstrap weighted multiple logistic regression procedure using SAS software. This method can be applied for the small sample size data especially where the data is very difficult to collect. By resampling (using bootstrap method), it provides the preliminary comprehensive information and also give the general overview on how the data behaviour even though the original data is not enough (small sample size). In our case, smaller standard error of the estimate parameter will tell us how accurate our estimate parameter is likely to be.
It is not easier to understand the behaviour of the data in studies when it is not reaching the actual sample size needed in an analysis [7].

REFERENCES


