International Journal of Engineering & Technology, 7 (4.30) (2018) 205-208

International Journal of Engineering & Technology

Website: www.sciencepubco.com/index.php/IJET

Research paper

Interaction Effects on Prediction of Children Weight at School Entry Using Model Averaging

Khuneswari Gopal Pillay[1]*, Sya Sya Syahira Muhammad Fitri Avtar[2] Mohd Asrul Affendi Abdullah[3]

[1,2,3]*Department of Mathemathics and Statistics, Faculty of Applied Science and Technology, Universiti Tun Hussein Onn Malaysia, Pagoh Campus, KM 1 Pachor Road,84000 Muar, Johor, Malaysia.*
**Corresponding author E-mail: khuneswari@uthm.edu.my*

Abstract

Model selection introduce uncertainty to the model building process, therefore model averaging was introduced as an alternative to overcome the problem of underestimate of standards error in model selection. This research also focused on using selection criteria between Corrected Akaike's Information Criteria (AIC_C) and Bayesian Information Criteria (BIC) as weight for model averaging when involving interaction effects. Mean squared error of prediction (MSE(P)) was used in order to determine the best model for model averaging. Gateshead Millennium Study (GMS) data on children weight used to illustrate the comparison between AIC_C and BIC. The results showed that model selection criterion AIC_C performs better than BIC when there are small sample and large number of parameters included in the model. The presence of interaction variable in the model is not significant compared to the main factor variables due to the lower coefficient value of interaction variables. In conclusion, interaction variables give less information to the model as it coefficient value is lower than main factor.

Keywords: *AIC_C; BIC; Interaction; Model Averaging; Model Selection.*

1. Introduction

A process of developing a probabilistic model that best describe the relationship between the dependent and independent variable is called model building. In model building, only the variable that best describe the model will be include. This process can be done by using several model building approaches. Model selection in practice requires the choice of a selection procedure, such as forward selection or backward elimination, coupled with a selection criterion, such as AIC or BIC, to select a small subset of variables to include in the model [2,8]. Model selection introduces additional uncertainty into the model-building process, but the standard errors of parameter estimates obtained from the selected model by standard statistical procedures will underestimate the true variability. The properties of standard parameter estimates obtained from the selected model do not reflect the stochastic nature of the model selection process [1].

In the literature [1,2,3,6], model averaging has been proposed as an alternative to model selection which is intended to overcome the underestimation of standard errors that is a consequence of model selection. Model averaging aims to incorporate the uncertainty associated with model selection into parameter estimation, by combining estimates over a set of possible models [12]. If the focus of model selection and model averaging is good prediction, then differences in the standard errors of estimators is not directly relevant to the comparison of these methods.

Model averaging is an alternative method for model selection. It allows to average the weights for a number of models, instead of picking one best model. Model averaging tends to shrink the estimates on the weaker terms, yielding better predictions. The “best” models will hold higher weights [1].

The effects of multicollinearity will be identified in model averaging when there exist interaction variables. BIC and AIC_c will be used to determine the best model in each of the model building approach. Lastly, mean squared error of prediction (MSE(P)) will be used to compare the model selection criteria when performing model averaging with interaction effects in term of prediction [3,4, 14].

This research illustrates the model-building approach using children weight at school entry and highlight the most significant factors contribute to children weight at school entry. The whole procedures of obtaining the best model will be explained step by step to provide a clear guideline of model-building approach on model averaging method.

2. Model-Building Approach

2.1. Multiple linear regression

Multiple linear regression (MR) is an extended version of simple linear regression model that involve two or more explanatory variables in a prediction equation. A more complex model that contain more explanatory variables typically is more useful in providing sufficiently precise of the response variable. The general MR model with more than two predictor variables is [9]

$$Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \ldots + \beta_{p-1} X_{i,p-1} + \varepsilon_i \quad (1)$$

where $\beta_0, \beta_1, \ldots, \beta_{p-1}$ are parameters, $X_{i1}, \ldots, X_{i,p-1}$ denotes the variable (which can be single independent variable, or interaction variable (first-order interaction, second-order interaction, third-order interaction, ...), or generated variable (polynomial variable, dummy variable) or transformed variables (ladder transformations, box-cox transformation, not smoothing or stationary)) and ε_i are independent $N(0,\sigma^2)$ for i = 1, ... , n. For example, the possible interaction variable is X_{12} (first-order interaction variables) which is the cross product of independent variables X_1 and X_2. Another example is the third-order interaction variable for three single independent variable (X_1, X_2 and X_3) is X_{123}.

2.2. Model selection criteria

Akaike's information criterion (AIC) is a popular and had been widely used as model selection criteria. AIC is calculated using the amount of fitted parameters, maximum likelihood estimate, and including intercept of model, *(p)*. AIC for model *M* is [2, 3, 13]

$$\text{AIC} = -2\ln \text{L(M)} + 2p \tag{2}$$

For small sample sizes (approximated as being when *n/p* is less than 40 where p is the number of fitted parameters and n is the sample size in the most complex model), a corrected version of Akaike's information criterion, (AIC_c) is recommended. The general form of AIC_c is [2, 3, 13]

$$\text{AIC}_c = AIC + \frac{2p\,(p+1)}{n-p-1} \tag{3}$$

Bayesian information criterion (BIC) is a criterion for model selection among a finite set of models. It is based, in part, on the likelihood function, and it is closely related to Akaike's information criterion (AIC). The BIC for model *M* is [2, 3]

$$\text{BIC} = -2\ln L(M) - (\log n)p \tag{4}$$

where *L(M)* is the maximized value of the likelihood function of model *M*, *n* is sample size of the data, *p* is the number of parameter in the model *M*.

2.3. Model averaging

Model selection is for including additional uncertainty into the model building process. The properties of parameter estimates obtained from the selected model method do not represent the stochastic nature of the model selection process. Model averaging had been proposed as an alternative method to model selection to overcome the under-estimation of standards errors in model selection. A model average estimator weighs across all possible models rather than picking a single best model. Model averaging will give less weight on the estimates of the weaker variables and will yield better predictions. The 'better' models will receive higher weights compare to others model. Suppose that there are *M* candidate models. In one approach, the weight w_m for model is [1, 2, 10, 12]

$$w_M = \frac{\exp\left(\frac{I_M}{2}\right)}{\sum_{m=1}^{M}\exp\left(\frac{I_M}{2}\right)} \tag{5}$$

where I_M is model selection criterion for m. The estimate of a parameter β_P is

$$\hat{\beta}_p = \sum_{m=1}^{M} w_M \hat{\beta}_{(p,M)} \tag{6}$$

where $\hat{\beta}_{(p,M)}$ is the estimate of $\hat{\beta}_p$ under model *M* for m = *1,2, …, M*. The modified weights will be used based on model selection criteria AIC, AIC_c and BIC. A modification was carried out for calculating the weights in order to avoid numerical error. The weights w_M were calculated as [8]

$$w_M = \frac{\exp\left(\frac{I_M-\bar{\iota}}{2}\right)}{\sum_{m=1}^{M}\exp\left(\frac{I_M-\bar{\iota}}{2}\right)} \tag{7}$$

where $\bar{\iota} = \frac{1}{M}\sum_{m=1}^{M} I_M$ with I_M is log-likelihood function of model *M* for m = *1, 2, …, M*.

2.4. Mean Square Error for prediction (MSE(P))

A reasonable measure for evaluating a model performance is by calculating its mean squared error of prediction (MSE(P)). In general MSE(P) can be describe as [3, 14]

$$MSE\ (P) = \frac{1}{t}\sum_{i=1}^{t}(\hat{y}_t - y_t)^2 \tag{8}$$

where $\hat{y}_t$ is estimated Y of test values and y_t is the actual test values used for prediction. MSE(P) is usually used to assess the performance of regressions.

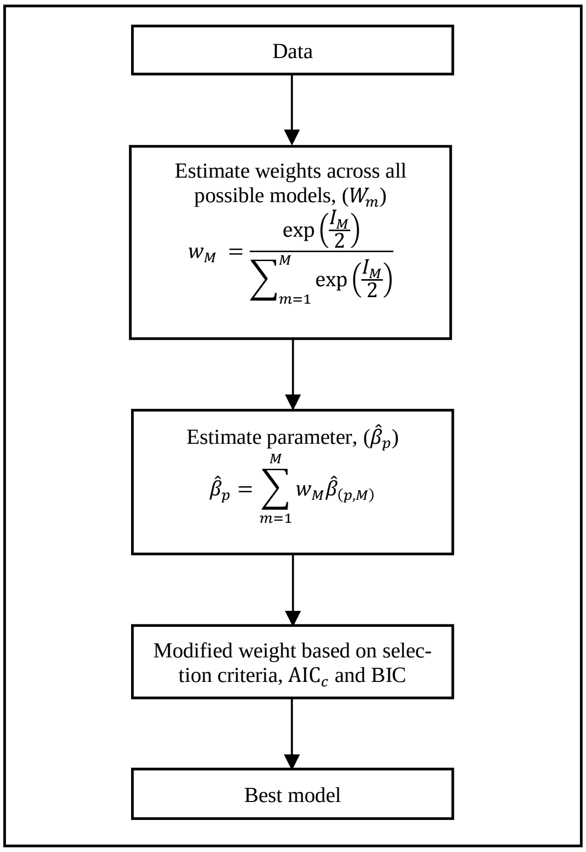

Fig 1: Framework of model averaging

3. Prediction Children Weight At School Entry

Study shows there are relationship between rapid weight gain and later overweight, leading to the suggestion that prevention and treatment of childhood obesity should begin as early as first year of life. The Gateshead Millennium Study (GMS) is a study of feeding and growth in infancy. The original objectives of the study are to explore the relationship between child development and feeding in the year of life, but it was later extended to follow up the children throughout childhood. The study was conducted at Gateshead area of northeast England. There are 1011 babies that born between June 1999 to 31 May 2000 that involve in this study. The sample size represents 83% of all births in the region on that year. The children were studied prospectively using parent report questionnaire shortly after birth at 6 weeks and at 4, 8, 12 months. The cohort has since been re-traced at school entry, parent report questionnaires completed at 5-8 years, and a range of anthropometric and body composition measures collected at age 7-8 years [8, 15].

Table 1: Description of variable for GMS

Variables	Descriptions	Unit
Y	Weight at school entry	Kilograms (kg)
X_1	Weight at 6 weeks	Kilograms (kg)
X_2	Weight at 4 months	Kilograms (kg)
X_3	Weight at 8 months	Kilograms (kg)
X_4	Weight at 12 months	Kilograms (kg)
X_5	Gender	1=Male, 2=Female

The analysis of male children and female children were carried out separately. Table 2 shows the descriptive statistics for male children. The mean weight of male children at school entry is 19.58kg while the median is 19.20kg. The first quartile is 17.80kg while third quartile is 21.00kg. It can be concluded that the smallest weight is 14.00kg while the heaviest male child weighted 34.60kg. 25% of the overall weight falls below 21.00kg and 25% of the overall weight is falls above 17.80kg. Table 3 shows the descriptive statistics for female children. The mean weight of female children at school entry is 19.90kg while the median is 19.00kg. The first quartile is 17.40kg while third quartile is 21.00kg. It can be concluded that the smallest weight is 13.00kg while the heaviest male child weighted 56.00kg. 25% of the overall weight falls below 21.00kg and 25% of the overall weight is falls above 56.00kg.

Table 2: Descriptive Statistics for male children

Statistics	Variables				
	Y	X_1	X_2	X_3	X_4
Minimum	14.00	3.21	4.50	6.80	7.54
Median	19.20	4.89	7.00	9.10	10.43
Mean	19.58	4.91	7.05	9.11	10.44
Maximum	34.60	6.61	9.75	13.28	14.30

Table 3: Descriptive Statistics for female children

Statistics	Variables				
	Y	X_1	X_2	X_3	X_4
Minimum	13.00	3.11	4.58	5.69	6.80
Median	19.00	4.57	6.33	8.34	9.58
Mean	19.90	4.58	6.42	8.38	9.80
Maximum	56.00	6.51	8.53	11.66	15.70

Figure 2 and 3 show the weights at school entry for male and female children, which are very similar on average. There are a number of exceptionally overweight children, especially female children.

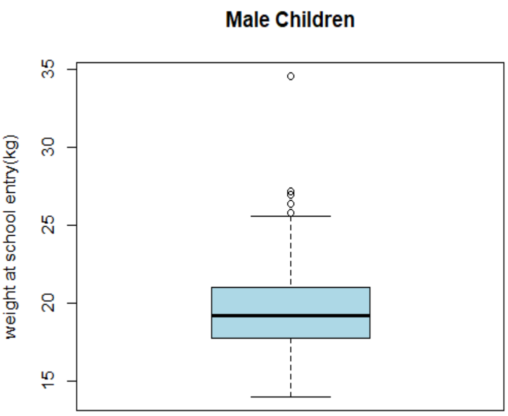

Fig 3: Boxplot of male children

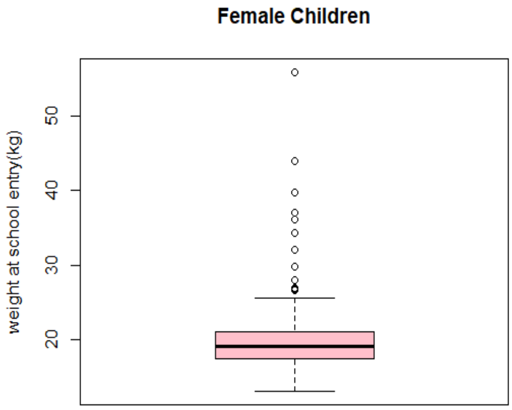

Fig 3: Boxplot of female children

Model averaging was carried out using two selection criteria as the weight for the models. The best model for model averaging will be choose based on the MSE(P) value of the model. Table 4 shows the coefficient and *p*-value of model averaging for male children, while Table 5 shows the output for female children. Coefficient for variable weight at 12 months (X_4) is significant in both outputs using BIC as model selection criteria in Table 4 and Table 5. This shows that there are significant effects of weight at 12 months on weight at school entry at age 5-8 years.

Table 4: Coefficient and *p*-value of model averaging for male children

Model selection criteria	AIC_c		BIC	
	Coefficient	*p*-value	Coefficient	*p*-value
Constant	0.658	0.125	0.4173	0
X_1	-7.562×10^{-2}	0.667	-2.491×10^{-4}	0.952
X_2	1.637×10^{-2}	0.865	-3.872×10^{-4}	0.919
X_3	-4.094×10^{-2}	0.428	-1.342×10^{-2}	0.131
X_4	-3.314×10^{-2}	0.430	-1.480×10^{-2}	0.019
X_1X_2	1.843×10^{-3}	0.841	-2.399×10^{-6}	0.990
X_1X_3	7.054×10^{-3}	0.694	-8.485×10^{-6}	0.980
X_1X_4	7.529×10^{-3}	0.674	1.226×10^{-5}	0.968
X_2X_3	-1.135×10^{-3}	0.902	3.586×10^{-5}	0.907
X_2X_4	-2.899×10^{-3}	0.758	1.768×10^{-5}	0.947
X_3X_4	2.860×10^{-3}	0.553	9.472×10^{-4}	0.223
$X_1X_2X_3$	-1.518×10^{-4}	0.857	-7.437×10^{-8}	0.996
$X_1X_2X_4$	-6.217×10^{-5}	0.926	-2.395×10^{-8}	0.997
$X_1X_3X_4$	-6.948×10^{-4}	0.701	-2.099×10^{-7}	0.991
$X_2X_3X_4$	2.853×10^{-4}	0.736	-8.517×10^{-9}	0.998
$X_1X_2X_3X_4$	-1.190×10^{-6}	0.977	-5.00×10^{-14}	1

Table 5: Coefficient and *p*-value of model averaging for female children

Model selection criteria	AIC_c		BIC	
	Coefficient	*p*-value	Coefficient	*p*-value
Constant	0.0139	0.019	0.009	0.001
X_1	-1.997×10^{-4}	0.871	-4.835×10^{-5}	0.838
X_2	-7.767×10^{-4}	0.495	-5.837×10^{-4}	0.730
X_3	-5.383×10^{-4}	0.503	-7.054×10^{-5}	0.750
X_4	-8.060×10^{-4}	0.169	-5.837×10^{-4}	0.018
X_1X_2	2.014×10^{-5}	0.905	2.541×10^{-7}	0.975
X_1X_3	1.046×10^{-5}	0.933	3.289×10^{-7}	0.967
X_1X_4	1.604×10^{-5}	0.875	3.728×10^{-6}	0.862
X_2X_3	5.966×10^{-5}	0.638	2.192×10^{-6}	0.899
X_2X_4	2.310×10^{-5}	0.783	7.961×10^{-6}	0.770
X_3X_4	2.010×10^{-5}	0.713	5.449×10^{-6}	0.775
$X_1X_2X_3$	-1.629×10^{-6}	0.910	-1.513×10^{-10}	0.999
$X_1X_2X_4$	-1.001×10^{-6}	0.930	-4.129×10^{-10}	0.998
$X_1X_3X_4$	-4.701×10^{-7}	0.947	-2.250×10^{-10}	0.999
$X_2X_3X_4$	-1.085×10^{-7}	0.980	-1.121×10^{-10}	0.999
$X_1X_2X_3X_4$	3.809×10^{-10}	0.998	1.032×10^{-18}	1

Table 6 shows the value of MSE(P) for all 4 models. Only one best model will be chosen for male and female children based on the lowest MSE(P) value. Based on Table 6, the best model for male and female children are the model obtained using AIC_c as selection criteria. AIC_c have lower MSE(P) due to the effect of small sample size.

Table 6: MSE(P) values

Model	Selection Criteria	MSE(P)
Male	AIC_c	$\mathbf{1.150 \times 10^{-4}}$
	BIC	1.340×10^{-4}
Female	AIC_c	$\mathbf{3.809 \times 10^{-7}}$
	BIC	9.443×10^{-6}

One of the assumption of regression is distribution of the residual should be normally distributed. In Figure 4, Q-Q plot is used to illustrate the distribution of residual for male children using AIC_c. The data seen to be normally distributed, however, Kolmogorov-Smirnov test is conducted to support the graphical method. Since the *p*-value of the Kolmogorov-Smirnov test is more than 0.05, so null hypothesis is rejected. The residuals are normally distributed.

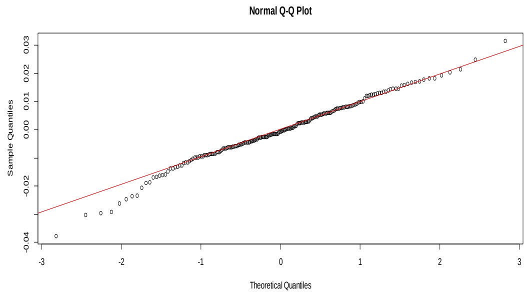

Fig 4: Q-Q plot for male children using AIC_c

Figure 5 shows that Q-Q of residuals for Female children using AIC_c. The data seen to be normally distributed, however, Kolmogorov-Smirnov test is conducted to support the graphical method. Since the p-value of the Kolmogorov-Smirnov test is more than 0.05, so null hypothesis is rejected. The residuals are normally distributed.

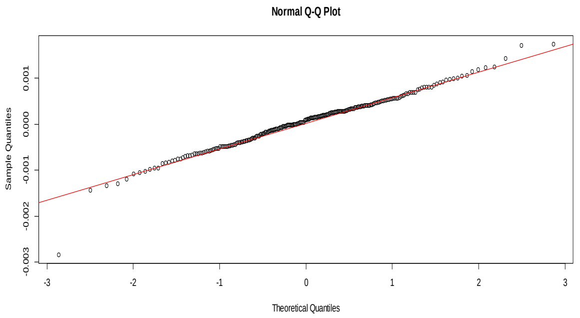

Fig 5: Q-Q plot for female children using AIC_c

4. Discussion and conclusion

Model-building using model averaging on children weight at school entry were illustrated clearly. This research considers up to high-order interaction to determine the best model. Model averaging being compare using AIC_C and BIC as selection criteria to modify the weight of parameter.
The results show that model selection criterion AIC_c, perform better than BIC for male children while for female children BIC is better compare to AIC_C. Theoretically AIC_c is known to be less biased than BIC when there is small sample size. However, there is only slightly different on the MSE(P) value of AIC_C and BIC for male and female children. As a conclusion, Model uncertainty is not an issued if the posterior concentrated on a single model and will lead to similar result for model averaging. Therefore, model selection criteria performance depends on the data [4].
The parameter with important information will be distribute more weight compare to the one with less information in model averaging method. Interaction variables will have smaller values of coefficient due to the cross-product of parameter, where the value of parameter getting larger and the coefficient will be smaller. Interaction variable is not significant compare to the main factor variables due to the lower coefficient value of interaction variables [8, 3, 9]. In conclusion, interaction variables give less information to the model as it coefficient value is lower than main factor.

5. Recommendation for Future Research

More predictor variable can be added to the model averaging in the future. However, the command for model averaging MuMIn package in R language only limits until 31 parameters, so it is advised to write a new command that can include more parameter in model averaging for R language in the future research [7].
Categorical variable or dummy variable can be introduced in the model. In this research, the variable gender for male and female children were analyses separately to form two different equations. It is suggested that the variable should be treated as dummy variable where 0 represent male and 1 represent female [5].
Other interaction variables also can be included in the model such as polynomial when it have quadratic response function or when the true curvilinear response function is unknown but a polynomial function is a good approximation to the true function [7].
As model averaging did not remove any of the variables from the model, a variable screening step can be used before deriving a model averaging predictor. A simple alternative screening procedure such as backward elimination which bases on the p-value can be used as a screening step to remove insignificant variables [11].

References

[1] Buckland TS, Burnham KP & Augustin NH (1997), Model selection: An integral part of inference. *Biometrics* 53(2), 603-618.
[2] Burnham KP & Anderson DR, *Model selection and multimodel inference: A Practical Information-Theoretic Approach*. New York: Springer-Verlag New York, Inc, (2002).
[3] Claeskens G & Hjort NL, *Model selection and model averaging*. Cambridge: University Press, (2008).
[4] Clyde M & Edward IG (2004), Model uncertainty. *Statistical Science* 19(1), 81-94.
[5] Hendry DF & Reade JJ (2005), Problems in model averaging with dummy variables, paper presented at Model Evaluation in Macroeconomics Workshop, University. of Oslo, Blindern, Norway.
[6] Hjort NL & Claeskens G (2003), Frequentist model average estimators. *Journal of the American Statistical Association* 98, 879-945.
[7] Kamil Barton (2016), MuMIn: Multi-Model Inference. R package version 1.15.6. Retrieved from URL. https://CRAN.R-project.org/package=MuMIn
[8] Khuneswari G, *Model selection and model averaging in the presence of missing data*. (Unpublished doctoral thesis). University of Glasgow, Scotland, United Kingdom, (2015).
[9] Kutner MH, Nachtsheim CJ & Neter J, *Applied Linear Regression Models*. New York: McGraw-Hill/Irwin. (2008).
[10] Mevik BH & Cederkvist HR (2004), Mean squared error of prediction (MSEP) estimates for principal component regression (PCR) and partial least squares regression (PLSR). *Journal of Chemometrics* 18(9), 422–429.
[11] Sauerbrei W, Hollander N & Buchholz A (2008). Investigation about a screening step in model selection. *Statistics and Computing* 18(2), 195-208.
[12] Schomaker M & Heumann C (2014), Model selection and model averaging after multiple imputation. *Computational Statistics and Data Analysis* 71, 758-770.
[13] Symonds MRE & Moussalli A (2011), A brief guide to model selection, multimodel inference and model averaging in behavioural ecology using Akaike's information criterion. *Behavioral Ecology and Sociobiology*, 65(1), 13–21.
[14] Wallach D & Goffinet B (1989), Mean squared error of prediction as a criterion for evaluating and comparing system models. *Ecological Modelling* 44, 299–306.
[15] Wright MW, Cox KM, Sherriff A, Franco-Villoria M, Pearce MS, Adamson AJ and core team GMS (2011), To what extent do weight gain and eating avidity during infancy predict later adposity? *Public Healthh Nutrition* 15(4), 656-662.