



Hierarchical Multiple Regression Model in Obtaining Body Weight Equation from Anthropometric Measurements

^{1*} H.J. Zainodin and ² S. J. Yap

¹*Faculty of Science and Natural Resources,
Universiti Malaysia Sabah,
88400 Kota Kinabalu, Sabah, Malaysia*

²*MAHSA College Sabah Campus,
Concourse Floor, 1Borneo Hypermall,
88400 Kota Kinabalu, Sabah, Malaysia*

E-mail: zainodin@gmail.com

*Corresponding author

ABSTRACT

Body weight has long been used as a parameter for calculation in many medical procedures. But when body weight is not available and its visual weight is inaccurate a body weight equation for both genders from anthropometric measurements is developed. This is the main aim of the study besides identifying the significance of interaction variables in hierarchically multiple regression analysis. The interaction variables involved here are up to the fifth-order (product of 6 independent variables). Here, the equation is developed using hierarchically multiple regression analysis. A modified method on the Zainodin-Noraini multicollinearity remedial method is proposed in this work to remedy the multicollinearity problem. Then, coefficient test is carried out on these models to eliminate insignificant variables from models that are free from multicollinearity problem. Ultimately, the equation developed in this work is free from multicollinearity problem and insignificant variables. The proposed modified method is found to be easier, time-saving more accurate and less prone to errors. An important finding in this work is that the interaction variables are significant and is best included in statistical analysis in order to yield a better prediction.

Keywords: Body weight equation, hierarchically multiple regression models, interaction variable, insignificant variable, multicollinearity effect, modified method.

1. INTRODUCTION

Chittawatanarat et al., 2012 stated that body weight is commonly used as one of the important clinical measurement parameters. Numerous clinical situations have used body weight as variable in determining the drug dosage, nutrition requirements, resuscitation process, pulmonary tidal volume estimation and haemodynamic assessments. However, the measurements of body weight are not always available and there are also some limitations to obtain body weight for certain people, especially the non ambulatory elderly people and critically ill patients. Although many body weight equations have been proposed, researchers seldom include interaction variable in their analyses. Thus, this work would develop a body weight equation from anthropometric measurements and identify the significance of interaction variables in multiple regression analysis. Apart from providing a route for those who are interested in developing body weight equations for different groups of people, the results of this work can also contribute to a better decision for researchers on including or excluding interaction variable from their analyses.

2. METHODOLOGY

Originally, there were 25 variables in the data set. Before starting the multiple regression analysis, the total number of all possible models has to be considered. The formula for counting number of all possible models can be found in Zainodin et al., 2011. Body weight, Y and the variable gender, G are exempted from the calculation of total number of all possible models because Y is the dependent variable of this work and G has been transformed into dummy variable in the original data set. Then, since the total number of all possible models of 23 variables is too massive at 96468992, not all the independent variables are used for analysis. For this reason, factor analysis and correlation coefficient values are used to identify suitable quantitative variables in this work.

In the factor analysis, G is again exempted from the variables identification because one of the aims of this work is to develop a body weight equation for both genders. Thus, only 24 variables are involved in the selection by using factor analysis. By suppressing the variables with absolute correlation values less than 0.75 in factor analysis, 11 out of 24 variables are found in the main cluster (the cluster where the dependent variable of this work belongs to). After excluding the dependent variable body weight from these 11 variables, 10 independent variables are left for further analysis.

Nevertheless, considering that the total number of all possible models for 10 single quantitative independent variables is 5120, which is quite large and hence, not preferred. Also, not all of the 10 single quantitative independent variables are used as independent variables in estimating body weight. Accordingly, to reduce the total number of all possible models, 6 out of 10 variables are chosen as the single quantitative independent variables, while the remaining 4 variables are transformed to single dummy variables. This is because the total number of all possible models for 6 single quantitative independent variables is 192 models and the number of parameters for a possible model (NP) for case with 6 single quantitative independent variables and 5 dummy variables (after adding the dummy variable gender, G) is 99, which are more possible to be carried out. The value of NP can be calculated by using the formula proposed by Zainodin *et al.* (2014). Thus, in this work, body weight (Y) is the dependent variable, chest diameter (X_1), shoulder girth (X_2), chest girth (X_3), bicep girth (X_4), forearm girth (X_5) and wrist girth (X_6) which have correlation coefficient values more than 0.8100 are selected as single quantitative independent variables and the ankle diameter (A), biacromial diameter (D), elbow diameter (B) and wrist diameter (R) which have correlation coefficient values less than 0.8100 are transformed to dummy variable. After variables identification, variable gender (G) is included in the work for further analysis. Then, multiple regression analysis is conducted on these identified variables by following the model building procedures proposed by Zainodin *et al.* (2011), as presented in Figure 1.

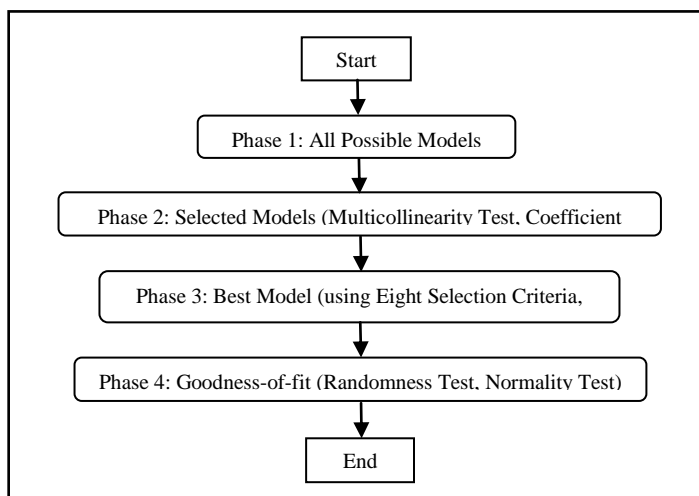


Figure 1: Model Building Procedures for Multiple Regressions

First, the total number of all possible models is counted and all of these models are listed out. This is based on the structural concept of Hierarchically Multiple Regression (Well-Formulated Models) as defined and illustrated in Jaccard, 2001. Then, by applying the concept of backward elimination, multicollinearity and coefficient tests are employed to discard variables systematically from each of the possible models. In the multicollinearity test, this work has utilised a modified method, which is different from the Zainodin-Noraini multicollinearity remedial method proposed in Zainodin et al., 2011. The removal steps of multicollinearity source variable by using modified method as shown in Figure 2. The details of the modified method are also discussed in Zainodin and Yap, 2013.

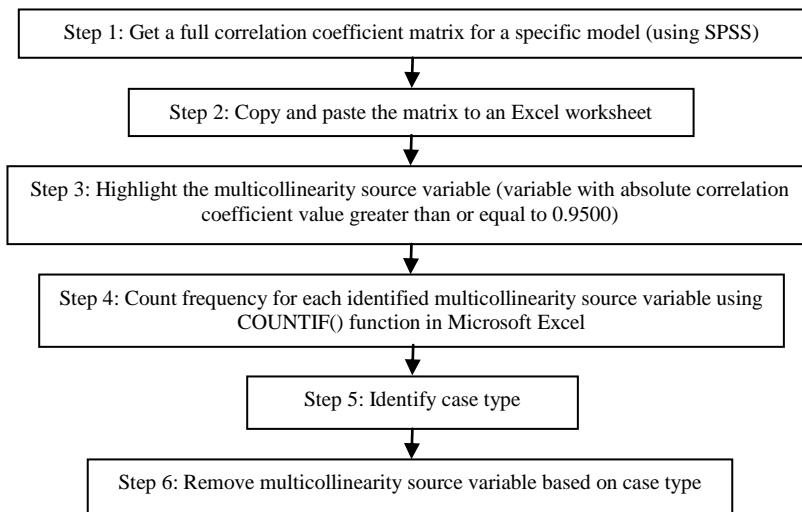


Figure 2: Multicollinearity Source Variable Removal Steps in Multicollinearity Test using Modified Method

According to Figure 2, a full correlation coefficient matrix is obtained by using SPSS. Then, this full correlation matrix is copied from SPSS and pasted in an Excel worksheet. After pasting the full matrix to an Excel worksheet, the variables with absolute correlation coefficient value greater than or equal to 0.9500 are highlighted. The value of 0.9500 is used because multicollinearity source variable is defined as variable with absolute correlation coefficient value greater than or equal to 0.9500 in this work. The choice of value in defining multicollinearity source variable is arbitrary. Researchers can use any values that are less or greater than 0.9500.

However, if researchers choose any values that are lower than 0.9500, and then they risked having too many variables removed from the models and vice versa. They may face the problem of having too many variables left in the model if they use any values that are greater than 0.9500. So after highlighting the multicollinearity source variables, the frequency of identified multicollinearity source variables in this specific model is counted using COUNTIF() function, which is different from the manual counting practised in the Z-N multicollinearity remedial method. If all of the correlation coefficient values between the independent variables are positive values, then the COUNTIF(range, "≥0.95") -1 is used. If there is negative correlation coefficient value between the independent variables, then the variable with correlation coefficient value less than -0.95 has to be considered. In this case, the COUNTIF(range, "≥0.95") +COUNTIF(range, "≤-0.95") -1 is used. After counting the frequencies for each identified multicollinearity source variable, the removal step is taken according to case type. The descriptions and removal steps for three different case types in the multicollinearity test are discussed in Zainodin and Yap, 2013. The removal steps are repeated until there is no more multicollinearity source variable left in the specific model.

3. ANALYSIS

From the 25 variables, six single quantitative independent variables and five single independent dummy variables are identified. So, by having six single quantitative independent variables, there are 192 possible models in this work. For convenience in getting the best model (or the body weight equation) out of the 192, they are grouped by category based on their highest order of interaction. Then, a best model is obtained from each of the category presented in Table 1.

Then, among the six best models obtained from these six categories, a best model will be obtained by using 8SC values. The interaction variables involved here are up to the fifth-order (product of 6 single independent quantitative variables). Based on Table 1, model with zero-order interaction falls into Category 1, model with first-order interaction to Category 2, that of second-order to Category 3. And finally, model with fifth-order interaction belongs to Category 6.

According to the model building procedures for multiple regression in Figure 1, in order to obtain selected models that are free from multicollinearity problem and insignificant variables from these 192 possible models, multicollinearity test and coefficient test have to be carried out on

each of these possible models. For simplicity an illustration of the procedures of these two tests, model M64 from Category 2 is considered and the same procedures are implemented on the remaining possible models. The equation of model M64 is presented as follows:

$$Y = f(X_1, X_2, X_{12}, D, B, R, A, G, X_1D, X_1B, X_1R, X_1A, X_1G, X_2D, X_2B, X_2R, X_2A, X_2G) \tag{1}$$

where X_{12} is a first-order interaction variable of chest diameter and shoulder girth, $X_1D, X_1B, X_1R, X_1A, X_1G$ are also first-order interaction variables of chest diameter and biacromial diameter, elbow diameter, wrist diameter, ankle diameter and gender, respectively and $X_2D, X_2B, X_2R, X_2A, X_2G$ are also first-order interaction variables of shoulder girth and biacromial diameter, elbow diameter, wrist diameter, ankle diameter and gender, respectively. This work provides a succinct summary of variables removed in the multicollinearity test, as presented in Table 2. Detailed procedures in removing multicollinearity source variables using modified method can be found in Zainodin and Yap, 2013.

TABLE 1: The Total Number of All Possible Models for Six Single Quantitative Independent Variables by Category

Number of Variables	Interactions						Total Models
	Zero-order	First-Order	Second-Order	Third-Order	Fourth-Order	Fifth-Order	
1	6	-	-	-	-	-	6
2	15	15	-	-	-	-	30
3	20	20	20	-	-	-	60
4	15	15	15	15	-	-	60
5	6	6	6	6	6	-	30
6	1	1	1	1	1	1	6
Total	63	57	42	22	7	1	192
Model Number	M1-M63	M64-M120	M121-M162	M163-M184	M185-M191	M192	
Category	1	2	3	4	5	6	

In Table 2, six variables are removed one by one systematically starting from model M64 then five variables are removed simultaneously from model M64.6. This is because model M64.6 belongs to Case C and there are five pairs of independent variables with highest frequency one. Thus, according to the removal steps for Case C as discussed in Zainodin and Yap, 2013, for each pair of variables with frequency one, independent variables that have weaker absolute correlation coefficient with dependent variable are removed from model M64.6 simultaneously.

TABLE 2: Summary of Variables Removed in the Multicollinearity Test

Models	Variables Removed	Resulting Models
M64	<i>B</i>	M64.1
M64.1	<i>A</i>	M64.2
M64.2	<i>D</i>	M64.3
M64.3	<i>G</i>	M64.4
M64.4	<i>R</i>	M64.5
M64.5	X_{12}	M64.6
M64.6	$X_1B, X_1A, X_2D, X_2R, X_2G$	M64.11

After removing 11 multicollinearity source variables from model M64, this model is now known as model M64.11 which is free from Multicollinearity effect. Then, according to Figure 1, coefficient test is carried out on model M64.11 to eliminate insignificant variables subsequently. An independent variable with highest p-value (>0.05) is eliminated and the resulting model is M64.11.1. Rerun the model and repeat the elimination procedures. Thus, the resulting model or selected model is model M64.11.2. The details of the coefficient test can be found in Zainodin et al., 2011 and Lind et al., 2012.

When all the tests (multicollinearity and coefficient) have been carried out on the 192 possible models, a list of selected models is obtained from each of the six categories. Then, from these six lists of selected models, a best model is obtained from each of them based on the 8SC. The best models from each category are shown in Table 3. This table also shows the number of parameters for each parent model (NP), error sum of squares (SSE)

and number of parameters for a selected model, $(k+1)$ for each category of these six best models. The model numbering used in this work is model $Ma.b.c$, where a denotes the number of the parent model, b represents the number of variables removed in multicollinearity test and c denotes the number of variables eliminated in the coefficient test.

TABLE 3: Best Models from Each Category

Categories	Best Models from Each Category	NP for parent model	SSE	$(k+1)$
1	M59.0.3: $Y = f(X_1, X_2, X_3, X_5, X_6, A, G)$	11	12099.0837	8
2	M116.34.3: $Y = f(X_1, X_2, X_3, X_5, X_6, X_{16}, X_{3A}, X_3G)$	46	11814.0038	9
3	M162.65.4: $Y = f(X_1, X_5, X_{36}, X_{123}, X_4D, X_4A, X_4G)$	77	11572.8726	8
4	M184.80.4: $Y = f(X_1, X_5, X_{36}, X_{123}, X_4D, X_4A, X_4G)$	92	11572.8726	8
5	M187.50.3: $Y = f(X_1, X_2, X_3, X_5, X_6, X_{16}, X_{3A}, X_3G)$	62	11814.0038	9
6	M191.86.4: $Y = f(X_1, X_5, X_{36}, X_{123}, X_4D, X_4A, X_4G)$	98	11572.8726	8

For instance, in Table 3, model M59.0.3 shows that zero or no variable is removed from the parent model, model M59 in the multicollinearity test and 3 variables are eliminated in the coefficient test. More details on the definition of the model numbering used in this work can be found in Zainodin et al., 2011. In Table 3, the value of $(k+1)$ can be calculated as follows:

$$(k + 1) = NP - b - c \tag{2}$$

For instance, the value of $(k+1)$ for model M116.34.3 can be calculated by using 46 to minus 34 and 3. The best models in Table 3 show that interaction variables are significant in the multiple regression analysis. This can be seen from the presence of interaction variables X_{16} , X_{36} , X_{123} , X_{3A} , X_3G , X_4D , X_4A and X_4G in the best model from each category, except category 1 which contains variables of zero-order interaction. In Table 3, the best model from each category is found to have the same independent variables, as presented in Table 4.

Hierarchical Multiple Regression Model in Obtaining Body Weight Equation from Anthropometric Measurements

TABLE 4: Best Models with the Same Independent Variables

Best Models with the Same Independent Variables	Remain	Remove
M116.34.3 = M187.50.3 $Y = f(X_1, X_2, X_3, X_5, X_6, X_{16}, X_{3A}, X_3G)$	M116.34.3	M187.50.3
M162.65.4 = M184.80.4 = M191.86.4: $Y = f(X_1, X_5, X_{36}, X_{123}, X_4D, X_4A, X_4G)$	M162.65.4	M184.80.4, M191.86.4

In Table 4, model M116.34.3 shares the same independent variables with model M187.50.3. Similarly, models M162.65.4, M184.80.4 and M191.86.4 have the same independent variables. Thus, only the lower order model (models M116.34.3 and M162.65.4) is used to represent the remaining models with the same independent variables. After identifying the representative models (removing the models that share the same independent variables) using 8SC value to determine the best model for this work. Detailed values are shown in Table 5.

Table 5: The Corresponding Selection Criteria Values for the Best Model

Category	Best Models	AIC	FPE	GCV	HQ	RICE	SCHWARZ	SGMASQ	SHIBATA
1	M59.0.3	25.94907	25.94915	25.95630	26.66598	25.96370	27.81233	25.52549	25.93509
2	M116.34.3	25.44301	25.44312	25.45200	26.23516	25.46122	27.50733	24.97675	25.42571
3	M162.65.4	24.82050	24.82058	24.82742	25.50623	24.83449	26.60272	24.41534	24.80713
4	M184.80.4	24.82050	24.82058	24.82742	25.50623	24.83449	26.60272	24.41534	24.80713
5	M187.50.3	25.44301	25.44312	25.45200	26.23516	25.46122	27.50733	24.97675	25.42571
6	M191.86.4	24.82050	24.82058	24.82742	25.50623	24.83449	26.60272	24.41534	24.80713

From Table 5, model M162.65.4 is found to be the best model since it has the most of the minimum values among the others in 8SC. Table 6 shows the coefficient table for the best model, model M162.65.4. Thus, by getting the best model, model M162.65.4, the estimated body weight can be calculated by using the following equation:

$$\hat{Y} = 4.1764 - 31.6248 * X_1 + 52.7577 * X_5 + 26.8483 * X_{36} + 69.9222 * X_{123} + 2.3012 * X_4D + 3.8301 * X_4A - 10.9476 * X_4G \tag{3}$$

As can be seen from Table 6, there are 8 estimated parameters (constant and 7 coefficients) left (i.e. $k + 1 = 8$) in the model M162.65.4. Then, from Table 3, total number of parameters in the parent model M162 is 77. Thus, using details procedure explained in Zainodin et al., 2014 it is confirmed that $(k + 1) = NP - b - c = 77 - 65 - 4 = 8$ where 65 coefficients are removed due to multicollinearity and 4 coefficients are eliminated due to insignificant effect. Similarly, we can confirm that the values obtained in the last column of Table 3 are correct and consistent.

After obtaining the best model, the goodness-of-fit is observed based on the residuals by carrying out the randomness test and normality test on the residuals.

TABLE 6: The Coefficient Values for the Best Model

Model M162.65.4	Unstandardized Coefficients		Standardized Coefficients	t_{cal}	p -value
	$\hat{\beta}$	Standard Error	Beta		
(Constant)	4.1764	7.0276		0.5943	5.5261×10^{-1}
X_1	-21.6248	10.9433	-0.1114	-1.9761	4.8726×10^{-2}
X_5	52.7577	9.3435	0.2802	5.6464	2.8259×10^{-8}
X_{36}	26.8483	8.5280	0.2339	3.1483	1.7461×10^{-3}
X_{123}	69.9222	9.5912	0.6574	7.2902	1.3056×10^{-12}
X_4D	2.3012	1.1639	0.0560	1.9771	4.8606×10^{-2}
X_4A	3.8301	1.0166	0.0970	3.7675	1.8565×10^{-4}
X_4G	-10.9476	1.3003	-0.2840	-8.4192	4.5348×10^{-16}

This is to ensure that the residuals meet the assumption of a multiple regression model. In this work, standardised residuals are used, rather than raw residuals. The null hypothesis of randomness test is H_0 : the standardised residuals are randomly distributed and the alternative hypothesis is H_1 : the standardised residuals are not randomly distributed.

Hierarchical Multiple Regression Model in Obtaining Body Weight Equation from Anthropometric Measurements

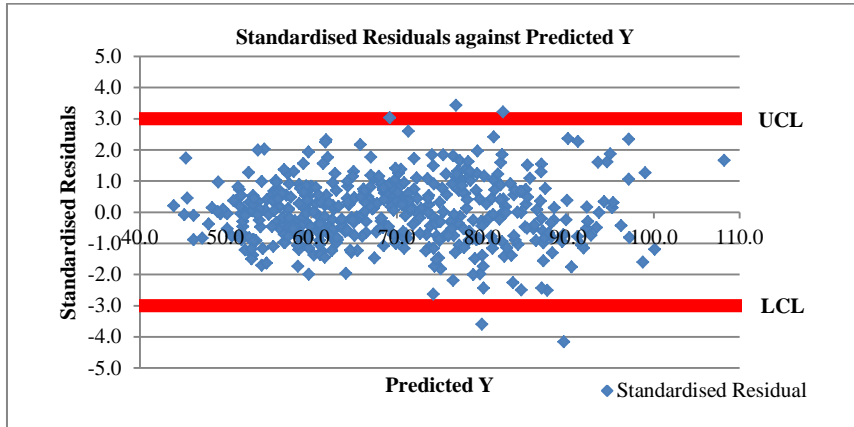


Figure 3: Scatter Plot of Standardised Residuals

Since there are 482 observations in the data set, a normal distribution table is used in the test. The null hypothesis is accepted when $|z_{cal}|$ is less than $z_{\alpha/2}$. The value of $|z_{cal}|$ is 0.0024 and the $z_{0.05}$ at ten percent level of significance is 1.645. Hence, the null hypothesis of randomness test is accepted, implying that the standardised residuals of the best model are randomly distributed.

TABLE 7: Kolmogorov-Smirnov test on Standardised Residuals

	Kolmogorov-Smirnov		
	Statistic	<i>df</i>	<i>p</i> -value
Standardised Residuals	0.0316	482	0.20000

Besides that, the scatter plot of the standardised residuals in Figure 3 also indicates that the standardised residuals are randomly distributed. This is because almost all (>98%) the standardised residuals in Figure 3 fall within the upper control limit (UCL) and lower control limit (LCL).

Then, the normality test is carried out to test the normality of the residuals. The null hypothesis of normality test is H_0 : the standardised residuals are normally distributed. The alternative hypothesis is H_1 : the standardised residuals are not normally distributed. The null hypothesis is rejected if the *p*-value is less than 0.05. The distribution of the residuals may

not show a standard normal distribution if the model chosen is incorrect. Table 7 shows the SPSS output of the Kolmogorov-Smirnov test on the standardised residuals. Since the p-value shown in Table 7 is 0.20 and is greater than 0.05, thus the quantitative null hypothesis of normality test is accepted, which implies that the standardised residuals are normally distributed. In addition, the histogram (supporting evidence) of the standardised residuals in Figure 4 showing a bell-shaped curve also implies that the standardised residuals follow a normal distribution. Thus, these show that the model chosen or the best model is acceptable and correct.

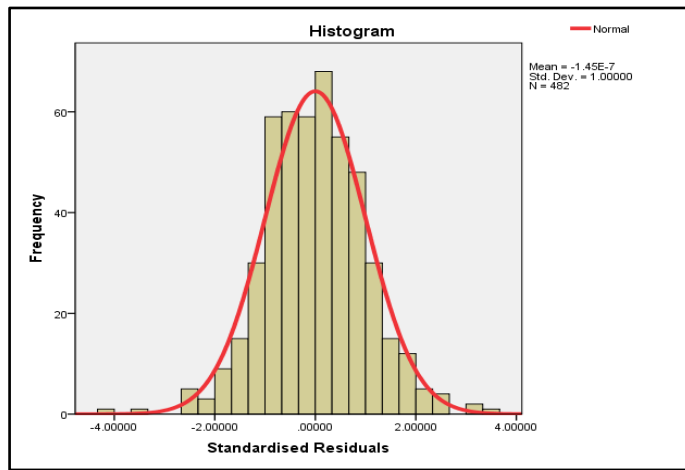


Figure 4: Histogram of Standardised Residuals

4. DISCUSSIONS

In equation (3), the partial regression coefficient of (X_1) is -21.6248. The negative sign on the coefficient denotes a negative relationship between the chest diameter (X_1) and the estimated body weight (\hat{Y}) . Each one unit increment on the chest diameter contributes to a 21.6248 times reduction on the estimated body weight, if all other variables in the best model are held constant. However, the estimated body weight has positive effect on X_{123} (interaction between (X_1, X_2, X_3)). This can be seen from the partial regression coefficient of X_{123} , which equals to 69.9222. The positive sign on the coefficient means that each one unit increment on X_{123} will result in a 69.9222 times increment on the estimated body weight if other variables in equation (3) are held constant. The partial regression coefficient of X_4D

(first-order interaction variable of bicep girth and transformed biacromial diameter) is 2.3012, which indicates that X_4D has positive effect on the estimated body weight \hat{Y} . Each unit increment on X_4D will result in an estimated 2.3012 times increment on the estimated body weight if other variables in equation (3) are held constant.

In this work, interaction variables are found to play an important role as the best model consisting of four first-order interaction variables (chest girth and wrist girth, and bicep girth each with biacromial diameter, ankle diameter, gender) and one second-order interaction variable (chest girth, chest diameter and shoulder girth). For better estimation it is suggested that interaction variables be included in an analysis. In general, the highest significant interaction order in the best model is about half of the highest order of interaction. This can be seen in the best model in equation (3), which is only up to the second-order although the analysis involved interaction variables up to the fifth-order. Similar results are also obtained by other researchers. In the work of Darmesah et al., 2008 on the number of asthmatics patients in Kota Kinabalu, Sabah, and interaction variables up to the third-order have been included. Nevertheless, the highest significant interaction variable left in the best model is only up to the first-order, which is about half of the highest order of interaction.

In 2009, Zainodin and Khuneswari, 2009 have developed interaction variables up to the fourth-order in their work on the house selling price model. However, the highest significant interaction variable found in the best model is only up to the second-order, which is also half of the highest order of interaction. Recent studies carried out by Zainodin et al., 2011 and Ansoorge et al., 2012, involved interaction variables up to the fourth- and seventh-orders respectively. Nevertheless, the highest significant interaction variable found in their best models are also about half of the highest order of interaction, that is, up to the first and fourth orders only respectively. Hence, it is quite safe to recommend that researchers include interaction variables up to half of the possible highest order of interaction (at least for analysis with 8 or less single independent quantitative variables). Convincing detailed evidence can be found in Yap, 2014.

5. CONCLUSION

Finally, a body weight equation, which is free from multicollinearity problem and insignificant variables, is developed. This body weight equation

offers a simple way in estimating body weight as the parameters used in equation (3) are common measurements of body dimensions. Apart from providing the steps in getting a body weight equation, the significance of interaction variables in multiple regression analysis is also identified in this work. The proposed modified method is easier, more time-saving, more accurate and less prone to human errors than the Zainodin-Noraini multicollinearity remedial method. Thus, the results of this work provide guidelines and references for other researchers who are interested in implementing similar studies.

ACKNOWLEDGEMENT

To facilitate the completion of this work we would like to express our special gratitude to the Sabah Government for Biasiswa Kerajaan Negeri Sabah (JPAN(B&L) 600-2007/399) partial support. Thanks to all concerned for their painstaking effort in careful reviewing, critical comments and meaningful suggestions in sprucing up this writing.

REFERENCES

- Ansorge, H. L., Adams, S., Jawad, A. F., Birk, D. E. and Soslowsky, L. J. (2012). Mechanical Property Changes during Neonatal Development and Healing using a Multiple Regression Model. *Journal of Biomechanics*, **45**(7): 1288-1292. DOI: 10.1016/j.jbiomech.2012.01.030.
- Chittawatanarat, K., Pruenglampoo, S., Trakulhoon, V., Ungpinitpong, W. and Patumanond, J. (2012). Development of Gender- and Age Group-Specific Equations for Estimating Body Weight from Anthropometric Measurement in Thai Adults. *International Journal of General Medicine*, **5**: 65-80. DOI: 10.2147/IJGM.S27507.
- Darmesah, G., Zainodin, H. J., Kamsia, B. and Suriani, H. (2008). Multiple Linear Regression in Forecasting the Number of Asthmatics. *WSEAS Transactions On Information Science & Applications*. **5**(6): 972-977.
- Jaccard, J. (2001). *Interaction Effects in Logistics Regression*. Thousand Oaks, California: SAGE Publications. DOI: <http://dx.doi.org/10.4135/9781412984515>

- Lind, D. A., Marchal, W. G. and Wathen, S. A. (2012). *Statistical Techniques in Business and Economics*. (15th edition). New York: McGraw-Hill Irwin.
- Zainodin, H. J. and Khuneswari, G. (2009). A Case Study on Determination of House Selling Price Model using Multiple Regression. *Malaysian Journal of Mathematical Sciences*. **3**(1): 27-44.
- Zainodin, H. J., Noraini, A. and Yap, S. J. (2011). An Alternative Multicollinearity Approach in Solving Multiple Regression Problems. *Trends in Applied Sciences Research*. **6**(11): 1241-1255. DOI: 10.3923/tasr.2011.
- Zainodin, H. J. and Yap, S. J. (2013). Overcoming Multicollinearity in Multiple Regression using Correlation Coefficient. *Proceedings of the International Conference on Mathematical Sciences and Statistics 2013. AIP Conference Proceedings*. **1557**: 416-419. DOI: 10.1063/1.4823947.
- Zainodin, H. J., Noraini Abdullah and Yap, S. J. (2014). Number of Parameters Counting in A Model. *Science International Journal*. **2**(2): 37-43 (ISSN2305-1884). DOI: 10.5567/sciint.2014.37.
- Yap, S. J. (2014). Modified method for removing multicollinearity problem in multiple regression models. Unpublished MSc UMS-thesis, Kota Kinabalu, Sabah: Universiti Malaysia Sabah.