

Number of Parameters Counting in a Hierarchically Multiple Regression Model

H.J. Zainodin, Noraini Abdullah and S.J. Yap

School of Science and Technology, Universiti Malaysia Sabah, Kota Kinabalu, Sabah, 88400, Malaysia

ABSTRACT

Background: Manually, when a dependent variable is affected by a large number of independent variables, the number of parameters helps researchers determine the number of independent variables to be considered in an analysis. However, when there are many parameters to be estimated in a model, the manual counting is tedious and time consuming. Thus, this study derives a method to determine the number of parameters systematically in a model. **Methods:** The model building procedure in this study involves removing variables due to multicollinearity and insignificant variables. Eventually, a selected model is obtained with significant variables. **Results:** The findings of this study would enable researchers to count the number of parameters in a resulting model (selected model) with ease and speed. On top of that, models which fulfill the assumption are considered in the statistical analysis. In addition, human errors caused by manual counting can also be minimised and avoided by implementing the proposed procedure. **Conclusion:** These findings will also undoubtedly help many researchers save time when their analyses involve complex iterations.

Key words: Number of parameters, single quantitative independent variable, dummy variable, interaction variable, hierarchically multiple regression model

Science International 2 (2): 37-43, 2014

INTRODUCTION

According to Ramanathan¹ in a linear regression model, regression coefficients are the unknown parameters to be estimated. In a simple linear regression, only two unknown parameters have to be estimated. However, problems arise in a multiple linear regression, when the numbers of parameters in the model are large and more complex, where three or more unknown parameters are to be estimated. Challenges arose when computer programme has to be written and complex iterations are required to perform with certain criterion. Thus, the exact number of parameters involved should be known in order to prepare the amount of data to suit such large and complex model. It is also important to note that in order to have a unique solution in finding the estimated parameters, according to the assumptions of multiple regression model stated by Gujarati and Porter², the number of estimated parameters must be less than the total number of observations.

According to Zainodin *et al.*³, Yahaya *et al.*⁴ in a multiple linear regression analysis, there are four phases in getting the best model, namely: Listing out all possible models, getting selected models, getting best model and conducting the validity of goodness-of-fit. In phase 1, the number of parameters for a possible model is denoted by

NP. To get the selected models, after listing out all of the possible models in phase 1, multicollinearity test and coefficient test are conducted on the possible models in phase 2. Before continuing to phase 2, number of parameters in each model must be less than the sample size, n . Discard the model which failed the initial criteria. In the multicollinearity test, multicollinearity source variables are removed from each of the possible models. Then, coefficient test is conducted on the possible models that are free from multicollinearity problem. Detailed procedure of this phase is explained in Zainodin *et al.*³. This is to eliminate insignificant variables from each of the possible models.

For a general model Ma.b.c with parent model number "a", the number of variables removed due to multicollinearity problem is denoted by b, the number of variables eliminated due to insignificance is denoted by $(k+1)$ and the resulting number of parameters for a selected model is represented by $(k+1)$. In most of the cases, if the number of the unknown parameters to be estimated for a possible model is large, then the number of parameters for a selected model will most probably be large too^{5,6}. In these cases, the manual counting on the large number of parameters is found to be time consuming. Furthermore, some of the parameters might be missed out due to human error in manual counting. Thus, the objective of this study is to propose a method

Corresponding Author: H.J. Zainodin, School of Science and Technology, Universiti Malaysia Sabah, Kota Kinabalu, Sabah, 88400, Malaysia

to count the number of parameters for a selected model, $(k + 1)$. The information of the NP, b and c is useful in getting the number of parameters for a selected model, $(k + 1)$.

According to Gujarati and Porter², a simple linear regression model without any interaction variable can be written as follows:

$$Y = \beta_0 + \beta_1 X_1 + u \tag{1}$$

where, Y is dependent variable, β_0 and β_1 are regression coefficients and they are the unknown parameters to be estimated, X_1 is single quantitative independent variable and u is error term. So, it can be observed that there are two unknown parameters to be estimated in Eq. 1. This equation can also be written in the form as follows:

$$Y = f(X_1)$$

Next, a hierarchically multiple linear regression models⁷ with interaction variable can be written as follows:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_{12} X_{12} + u \tag{2}$$

where, Y is dependent variable, β_0 , β_1 , β_2 and β_{12} are regression coefficients and they are the unknown parameters to be estimated, X_1 and X_2 are single quantitative independent variables, X_{12} is first-order interaction variable and u is error term. Thus, it can be seen that there are four unknown parameters to be estimated in Eq. 2. This equation can also be written in the following form:

$$Y = f(X_1, X_2, X_{12})$$

Next, an example for a linear regression model with interaction variables and dummy variables is shown as follows:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_{12} X_{12} + \beta_D D + \beta_{1D} X_1 D + \beta_{2D} X_2 D + u \tag{3}$$

where, Y is dependent variable, β_0 , β_1 , β_2 , β_{12} , β_D , β_{1D} and β_{2D} are regression coefficients and they are the unknown

parameters to be estimated, X_1 and X_2 are single quantitative independent variables, X_{12} is first-order interaction variable, D is single independent dummy variable, $X_1 D$ is first-order interaction variable of X_1 and D, $X_2 D$ is first-order interaction variable of X_2 and D and u is error term. Therefore, it can be observed that there are seven unknown parameters to be estimated in Eq. 3. This equation can also be written as follows:

$$Y = f(X_1, X_2, X_{12}, D, X_1 D, X_2 D)$$

Equation 1-3 can be written in general model in the form:

$$Y = S_0 + S_1 W_1 + S_2 W_2 + \dots + S_k W_k + u \tag{4}$$

where, S_0 is the intercept and S_j is the jth partial regression coefficient of the corresponding independent variable W_j for $j = 1, 2, \dots, k$.

According to Zainodin *et al.*³, independent variable W_j included the single independent variables, interaction variables, generated variables, dummy variables and transformed variables. In this study, $(k + 1)$ denotes number of parameters for a selected model. The corresponding labels of the general model in Eq. 4-3 are shown in Table 1.

From Table 1, it is known that β_0 represents the S_0 in the general model, β_1 represents the S_1 and the same goes to other estimated parameters in Table 1. Variable X_1 in Eq. 3 represents W_1 in the general model and the same goes to other variables in Table 1.

Instead of counting the number of parameters one by one as above, a equation to count the number of parameters in model without interaction variable and in model with interaction variable is proposed in this study. The Eq. 5 is presented as follows:

$$NP = \begin{cases} g + h + 1, & v = 0 \\ \sum_{i=1}^{v+1} C_i + (g + 1)h + 1, & v = 1, 2, \dots \end{cases} \tag{5}$$

where, NP is number of parameters for a possible model, g is number of single independent quantitative variables, h is number of single independent dummy variables and v is highest order of interaction (between single

Table 1: Corresponding labels of general model and Eq. 3

No. of independent variables (k)	Coefficients		Independent variables	
	General model	Eq. 3	General model	Eq. 3
0	S_0	β_0	-	-
1	S_1	β_1	W_1	X_1
2	S_2	β_2	W_2	X_2
3	S_3	β_{12}	W_3	X_{12}
4	S_4	β_D	W_4	D
5	S_5	β_{1D}	W_5	$X_1 D$
6	S_6	β_{2D}	W_6	$X_2 D$

independent quantitative variable) in the model. Here, $v = 0$ denotes model without interaction variable (or model with zero-order interaction variable) and $v = 1, 2, \dots$ denotes model with first or higher order interaction variable(s). Hence, Eq. 5 can now be tested in the following instances to prove its validity in counting the number of parameters in a hierarchically multiple regression model. The aim of this study is to propose a method to count the number of parameters for a selected model, $(k+1)$. The information of the NP, b and c is useful in getting the number of parameters for a selected model, $(k+1)$.

MATERIALS AND METHODS

This will help to better understand the application of the equation as mentioned in previous section. In order to achieve a model free from multicollinearity effects and insignificant effects the following 4 phase model building procedure is implemented (details can be found in^{3,4}).

- C **Phase 1:** All possible models
- C **Phase 2:** Selected model
Multicollinearity test and coefficient test (Include NPM is Near Perfect Multicollinearity test and NPC is Near Perfect Collinearity test)
- C **Phase 3:** Best model
- C **Phase 4:** Goodness-of-fit

Randomness test and normality test: This study also revealed the parameters of independent variables measure multicollinearity effect between independent variables parameters and structural parameters. As discussed in earlier section, the number of parameters is very important before arriving at a selected and best model. Thus, some illustrations follow:

Models without interaction variable: Here, the following models are considered without interaction variable or with zero-order interaction variable. As pointed out earlier in Eq. 5, the number of parameters in the case of model without interaction variable (or $v = 0$) can be computed using $g+h+1$. For instance, consider Eq. 6, a model with zero-order interaction variable (or v equals to 0), number of single quantitative independent variable, g equals to 1 and number of single dummy variable, h equals to 5, as follows:

$$Y = f(X_1, D, B, R, A, G) \quad (6)$$

where, Y is a dependent variable, X_1 is a single quantitative independent variable ($g = 1$) and D, B, R, A and G are 5 single dummy variables ($h = 5$). Then, total number of parameters involved is:

$$NP = g+h+1 = 7$$

For simplicity, consider another example for calculating the number of parameters in a model without an interaction variable. Considering Eq. 7 which is a model with zero-order interaction variable (or v equals to 0), number of single quantitative independent variable, g equals to 8 and number of single dummy variable, h equals to 10, as follows:

$$Y = f(X_1, X_2, X_3, X_4, X_5, X_6, X_7, X_8, B, C, L, E, W, K, A, G, H, S) \quad (7)$$

where, $X_1, X_2, X_3, X_4, X_5, X_6, X_7$ and X_8 are single quantitative independent variables and $B, C, L, E, W, K, A, G, H$ and S are 10 single dummy variables. Then, the following is obtained:

$$NP = g+h+1 = 19$$

Models with interaction variable: Now, consider models with interaction variable (i.e., $v = 1, 2, \dots$) in this subsection. According to the equation mentioned in Eq. 5, it is known that the number of parameters in a model with interaction variable is calculated in a different way from a model without interaction variable. A few examples are presented to provide better understanding of this equation. For instance, model with interaction variable up to first-order (i.e., v equals to 1), number of single quantitative independent variable, g equals to 2 and number of single dummy variable, h equals to 5 is presented in Eq. 8:

$$Y = f(X_2, X_4, X_{24}, D, B, R, A, G, X_2D, X_2B, X_2R, X_2A, X_2G, X_4D, X_4B, X_4R, X_4A, X_4G) \quad (8)$$

where, X_2 and X_4 are single quantitative independent variables, D, B, R, A and G are single dummy variables and $X_{24}, X_2D, X_2B, X_2R, X_2A, X_2G, X_4D, X_4B, X_4R, X_4A$ and X_4G are first-order interaction variables. Then, this led to:

$$\begin{aligned} NP &= \sum_{i=1}^{v+1} {}^g C_i + [(g+1)h+1] \\ &= \sum_{i=1}^{1+1} {}^2 C_i + [(2+1)5+1] \\ &= ({}^2 C_1 + {}^2 C_2) + 16 \\ &= ({}^2 C_1 + {}^2 C_2) + 16 \\ &= 19 \text{ (Total No. of parameters as in Eq. 8)} \end{aligned}$$

Next, consider a larger model with higher order of interaction variable, for instance, a model with fifth-order interaction (i.e., v equals to 5), number of

single quantitative independent variable, g equals to 6 and number of single dummy variable, h equals to 5 as presented in Eq. 9:

$$Y = f(X_1, X_2, X_3, X_4, X_5, X_6, X_{12}, X_{13}, X_{14}, X_{15}, X_{16}, X_{23}, X_{24}, X_{25}, X_{26}, X_{34}, X_{35}, X_{36}, X_{45}, X_{46}, X_{56}, X_{123}, X_{124}, X_{125}, X_{126}, X_{134}, X_{135}, X_{136}, X_{145}, X_{146}, X_{156}, X_{234}, X_{235}, X_{236}, X_{245}, X_{246}, X_{256}, X_{345}, X_{346}, X_{356}, X_{456}, X_{1234}, X_{1235}, X_{1236}, X_{1245}, X_{1246}, X_{1256}, X_{1345}, X_{1346}, X_{1356}, X_{1456}, X_{2345}, X_{2346}, X_{2356}, X_{2456}, X_{3456}, X_{12345}, X_{12346}, X_{12356}, X_{12456}, X_{13456}, X_{23456}, X_{123456}, D, B, R, A, G, X_1D, X_1B, X_1R, X_1A, X_1G, X_2D, X_2B, X_2R, X_2A, X_2G, X_3D, X_3B, X_3R, X_3A, X_3G, X_4D, X_4B, X_4R, X_4A, X_4G, X_5D, X_5B, X_5R, X_5A, X_5G, X_6D, X_6B, X_6R, X_6A, X_6G) \tag{9}$$

In Eq. 9, X_1, X_2, X_3, X_4, X_5 and X_6 are single quantitative independent variables, $X_{12}, X_{13}, X_{14}, X_{15}, X_{16}, X_{23}, X_{24}, X_{25}, X_{26}, X_{34}, X_{35}, X_{36}, X_{45}, X_{46}, X_{56}, X_{1D}, X_{1B}, X_{1R}, X_{1A}, X_{1G}, X_{2D}, X_{2B}, X_{2R}, X_{2A}, X_{2G}, X_{3D}, X_{3B}, X_{3R}, X_{3A}, X_{3G}, X_{4D}, X_{4B}, X_{4R}, X_{4A}, X_{4G}, X_{5D}, X_{5B}, X_{5R}, X_{5A}, X_{5G}, X_{6D}, X_{6B}, X_{6R}, X_{6A}$ and X_{6G} are first-order interaction variables, $X_{123}, X_{124}, X_{125}, X_{126}, X_{134}, X_{135}, X_{136}, X_{145}, X_{146}, X_{156}, X_{234}, X_{235}, X_{236}, X_{245}, X_{246}, X_{256}, X_{345}, X_{346}, X_{356}$ and X_{456} are second-order interaction variables, $X_{1234}, X_{1235}, X_{1236}, X_{1245}, X_{1246}, X_{1256}, X_{1345}, X_{1346}, X_{1356}, X_{1456}, X_{2345}, X_{2346}, X_{2356}, X_{2456}$ and X_{3456} are third-order interaction variables, $X_{12345}, X_{12346}, X_{12356}, X_{12456}, X_{13456}$ and X_{23456} are fourth-order interaction variables and X_{123456} is fifth-order interaction variable. Then, total number of parameters:

$$\begin{aligned} NP &= \sum_{i=1}^{v+1} {}^g C_i + [(g+1)h+1] \\ &= \sum_{i=1}^{5+1} {}^6 C_i + [(6+1)5+1] \\ &= \left(\sum_{i=1}^6 {}^6 C_i\right) + 36 \\ &= 99 \end{aligned}$$

Lastly, consider another example of a larger model which has interaction variable up to 6th order and nine single dummy variables. Consider Eq. 1, a model with highest order of interaction, v equals to 6, number of single quantitative independent variable, g equals to 7 and number of single dummy variable, h equals to 9.

$$Y = f(X_1, X_2, X_3, X_4, X_5, X_6, X_7, X_{12}, X_{13}, X_{14}, X_{15}, X_{16}, X_{17}, X_{23}, X_{24}, X_{25}, X_{26}, X_{27}, X_{34}, X_{35}, X_{36}, X_{37}, X_{45}, X_{46}, X_{47}, X_{56}, X_{57}, X_{67}, X_{123}, X_{124}, X_{125}, X_{126}, X_{127}, X_{134}, X_{135}, X_{136}, X_{137}, X_{145}, X_{146}, X_{147}, X_{156}, X_{157}, X_{167}, X_{234}, X_{235}, X_{236}, X_{237}, X_{245}, X_{246}, X_{247}, X_{256}, X_{257}, X_{267}, X_{345}, X_{346}, X_{347}, X_{356}, X_{357}, X_{367}, X_{456}, X_{457}, X_{467}, X_{567}, X_{1234}, X_{1235}, X_{1236}, X_{1237}, X_{1245}, X_{1246}, X_{1247}, X_{1256}, X_{1257}, X_{1267}, X_{1345}, X_{1346}, X_{1347}, X_{1356}, X_{1357}, X_{1367}, X_{1456}, X_{1457}, X_{1467}, X_{1567}, X_{2345}, X_{2346}, X_{2347}, X_{2356}, X_{2357}, X_{2367}, X_{2456}, X_{2457}, X_{2467}, X_{2567}, X_{3456}, X_{3457}, X_{3467}, X_{3567}, X_{4567}, X_{12345}, X_{12346}, X_{12347}, X_{12356}, X_{12357}, X_{12367}, X_{12456}, X_{12457}, X_{12467}, X_{12567}, X_{13456}, X_{13457}, X_{13467}, X_{13567}, X_{14567}, X_{23456}, X_{23457}, X_{23467}, X_{23567}, X_{24567}, X_{34567}, X_{123456}, X_{123457}, X_{123467}, X_{123567}, X_{134567}, X_{234567}, X_{1234567}, B, C, E, W, K, A, G, H, S, X_1B, X_1C, X_1E, X_1W, X_1K, X_1A, X_1G, X_1H, X_1S, X_2B, X_2C, X_2E, X_2W, X_2K, X_2A, X_2G, X_2H, X_2S, X_2B, X_3C, X_3E, X_3W, X_3K, X_3A, X_3G, X_3H, X_3S, X_3B, X_3C, X_4E, X_4W, X_4K, X_4A, X_4G, X_4H, X_4S, X_5B, X_5C, X_5E, X_5W, X_5K, X_5A, X_5G, X_5H, X_5S, X_6B, X_6C, X_6E, X_6W, X_6K, X_6A, X_6G, X_6H, X_6S, X_7B, X_7C, X_7E, X_7W, X_7K, X_7A, X_7G, X_7H, X_7S) \tag{10}$$

Here, $X_{123456}, X_{123457}, X_{123467}, X_{123567}, X_{124567}, X_{134567}$ and X_{234567} are 5th order interaction variables and $X_{1234567}$ is a 6th order interaction variable. Then, total number of parameters:

$$\begin{aligned} NP &= \sum_{i=1}^{v+1} {}^g C_i + [(g+1)h+1] \\ &= \left(\sum_{i=1}^{6+1} {}^7 C_i\right) + 73 \\ &= ({}^7 C_1 + {}^7 C_2 + {}^7 C_3 + {}^7 C_4 + {}^7 C_5 + {}^7 C_6 + {}^7 C_7) + 73 \\ &= 200 \end{aligned} \tag{11}$$

As can be seen from the illustrations, number of parameters calculated using the derived equation tally with that in manual counting.

RESULTS AND DISCUSSION

The proposed equation defined in Eq. 5 is especially useful in counter checking the variables when listing all of the possible models in an analysis. This is because some of the variables might be missed out when there are a large number of parameters involved in a possible model. Table 2 shows all the possible models for an

Table 2: No. of parameters in each of all possible models

All possible models	NP	g	h	v
M1: $Y = f(X_1, D)$	3	1	1	0
M2: $Y = f(X_2, D)$	3	1	1	0
M3: $Y = f(X_3, D)$	3	1	1	0
M4: $Y = f(X_1, X_2, D)$	4	2	1	0
M5: $Y = f(X_1, X_3, D)$	4	2	1	0
M6: $Y = f(X_2, X_3, D)$	4	2	1	0
M7: $Y = f(X_1, X_2, X_3, D)$	5	3	1	0
M8: $Y = f(X_1, X_2, X_{12}, D, X_1D, X_2D)$	7	2	1	1
M9: $Y = f(X_1, X_3, X_{13}, D, X_1D, X_3D)$	7	2	1	1
M10: $Y = f(X_2, X_3, X_{23}, D, X_2D, X_3D)$	7	2	1	1
M11: $Y = f(X_1, X_2, X_3, X_{12}, X_{13}, X_{23}, D, X_1D, X_2D, X_3D)$	11	3	1	1
M12: $Y = f(X_1, X_2, X_3, X_{12}, X_{13}, X_{23}, X_{123}, D, X_1D, X_2D, X_3D)$	12	3	1	2

analysis that has two single quantitative independent variables (X_1 and X_2) and one single dummy variable (D).

In Table 2, with the information on g , h and v , the NP for possible models M1-M12 can be computed by using Eq. 5. Then, the number of parameters for each of the possible models can be counterchecked by using the computed NP values. For simplicity, each of the 12 models can be written in general form as in Eq. 4.

After introducing the equation in counting the number of parameters for a possible model, the way of getting the number of parameters for a selected model, $(k+1)$ is presented^{8,9}. A model M32, is used as an illustration and it had also been mentioned earlier in Eq. 3. So, multicollinearity test is conducted on the possible model, model M32. The removal of multicollinearity source variables from this model are shown in Table 2-4. This study uses the modified method in removing multicollinearity source variables (excel command: COUNTIF()).

In Table 3, all of the multicollinearity source variables (variables with absolute correlation coefficient values greater than or equal to 0.9500) are circled. Then, it is found that variables X_{12} , D , X_1D and X_2D have frequencies 2. So, according to³, model M32 belongs to case B. To avoid confusion between dummy variable B and case B, case B is represented by case 2 in this study.

Similarly, case C is represented by case 3 in this study. Based on the removal steps for case 2, variable X_{12} which has the weakest absolute correlation coefficient with dependent variable Y , if compared to variables D , X_1D and X_2D , is removed from model M32. The same removal steps are carried out on the reduced model; model M32.1, as presented in Table 4.

After removing variable X_2D from model M32.1 in Table 4, details correlation coefficients of the reduced model M32.2 are shown in Table 5. It is observed that each of the variables D and X_1D has the highest frequency of one, respectively. So, it is identified that model M32.2 belongs to case III. Therefore, variable X_1D which has a weaker absolute correlation coefficient with the dependent variable Y is then removed from this model. Thus, the resulting model free from multicollinearity is M32.3.

Table 6 shows that model M32.3 is free from multicollinearity source variables because all the absolute correlation coefficient values between all the independent variables are less than 0.9500 (except the diagonal values).

By observing the model M32.3, it is found that the number of variables removed due to multicollinearity problem is 3. Therefore, b for model M32.3 is 3. More details on the definition of model name can be found in³ and^{4,6,10}. Thus the resulting model, M32.3 is free from multicollinearity effects and the coefficient test is conducted on the model. The task then is to eliminate insignificant variables from this model, M32.3.

In Table 7, it is found that variable X_2 has the highest p -value among other independent variables and is greater than 0.05 (since the number of single quantitative independent variables is greater than 5 and the coefficient test is a two-tail, the level of significance is set at 10%. This is based on¹¹ recommendations). Thus, variable X_2 is eliminated from model M32.3 and this reduced model is called model M32.3.1 as 1 variable is eliminated due to insignificance. Details on the coefficient test can be found in^{8,12,13}.

Table 8 shows that model M32.3.1 is free from insignificant variable because both the p -values of variable X_1 and D are less than 0.05. From Table 8, it is found that 3 parameters (constant of the model, coefficient of variables X_1 and D) are left in model M32.3.1, or in other words, $(k+1)$ equals to 3. From the model name, model M32.3.1; it is noticed that 1 variable is eliminated in coefficient test and c equals to 1. As mentioned earlier in Eq. 3, there are seven unknown parameters to be estimated for model M32, so NP equals to 7. Thus, by knowing the NP, b and c for model M32.3.1 (i.e., Ma.b.c), the number of parameters $(k+1)$ for model M32.3.1 can be counterchecked using the proposed equation in this study as:

$$\begin{aligned}(k+1) &= NP-b-c \\ &= 7-3-1 \\ &= 3\end{aligned}\tag{12}$$

Therefore, it is shown in Table 8 that the number of parameters left in the selected model M32.3.1 is the same

Table 3: Process of removing a multicollinearity source variable from model M32

Parameters	Y	X_1	X_2	X_{12}	D	X_1D	X_2D
Y	1	0.8134	0.7275	0.7505	-0.8081	-0.7883	-0.7856
X_1	0.8134	1	0.9239	0.9611	-0.6331	-0.5954	-0.5930
X_2	0.7275	0.9239	1	0.9828	-0.5320	-0.5007	-0.4844
X_{12}	0.7505	0.9611	0.9828	1	-0.5483	-0.5186	-0.5098
D	-0.8081	-0.6331	-0.5320	-0.5483	1	0.9955	0.9930
X_1D	-0.7883	-0.5954	-0.5007	-0.5186	0.9955	1	0.9977
X_2D	-0.7856	-0.5930	-0.4844	-0.5098	0.9930	0.9977	1
Frequency		1	1	2	2	2	2
Case	II						
Remove	X_{12}						

Table 4: Process of removing a multicollinearity source variable from model M32.1

Parameters	Y	X ₁	X ₂	D	X ₁ D	X ₂ D
Y	1	0.8134	0.7275	-0.8081	-0.7883	-0.7856
X ₁	0.8134	1	0.9239	-0.6331	-0.5954	-0.5930
X ₂	0.7275	0.9239	1	-0.5320	-0.5007	-0.4844
D	-0.8081	-0.6331	-0.5320	1	0.9955	0.9930
X ₁ D	-0.7883	-0.5954	-0.5007	0.9955	1	0.9977
X ₂ D	-0.7856	-0.5930	-0.4844	0.9930	0.9977	1
Frequency				2	2	2
Case	II					
Remove	X ₂ D					

Table 5: Process of removing a multicollinearity source variable from model M32.2

Parameters	Y	X ₁	X ₂	D	X ₁ D
Y	1	0.8134	0.7275	-0.8081	-0.7883
X ₁	0.8134	1	0.9239	-0.6331	-0.5954
X ₂	0.7275	0.9239	1	-0.5320	-0.5007
D	-0.8081	-0.6331	-0.5320	1	0.9955
X ₁ D	-0.7883	-0.5954	-0.5007	0.9955	1
Frequency				1	1
Case	III				
Remove	X ₁ D				

Table 6: Correlation coefficient matrix of model M32.3 which is free from multicollinearity source variable

Parameters	Y	X ₁	X ₂	D
Y	1	0.8134	0.7275	-0.8081
X ₁	0.8134	1	0.9239	-0.6331
X ₂	0.7275	0.9239	1	-0.5320
D	-0.8081	-0.6331	-0.5320	1

Table 7: Elimination of insignificant variable from model M32.3

Model M32.3	Unstandardized coefficients		Standardized coefficient			Action
	$\hat{\beta}$	SE	Beta	t _{cal}	p-value	
Constant	-13.5560	2.9287		-4.6286	5.9385 × 10 ⁶	
X ₁	0.3810	0.0632	0.4909	6.0262	6.0366 × 10 ⁹	
X ₂	0.0302	0.1708	0.0131	0.1766	8.5999 × 10 ¹	Eliminate
D	-8.3701	0.6285	-0.4903	-13.3183	6.8515 × 10 ³¹	

Table 8: Model M32.3.1 which is free from Insignificant variable

Model M32.3.1	Unstandardized coefficients		Standardized coefficient			Action
	$\hat{\beta}$	SE	Beta	t _{cal}	p-value	
Constant	-13.7221	2.7682		-4.9570	1.3247 × 10 ⁶	End
X ₁	0.3910	0.0281	0.5038	13.9352	5.0915 × 10 ³³	
D	-8.3503	0.6171	-0.4891	-13.5305	1.2226 × 10 ³¹	

with the value obtained from the proposed in Eq. 11. In line with the above discussion, other researchers also highlighted the importance of this parameters counting. They ranked them (as importance, significance or dependency etc.) the parameters of a model based on the magnitude of the coefficients^{11,14,15}.

CONCLUSION

This study is new and groundbreaking. It has succeeded in proposing a equation in counting the number of parameters for each of all possible models (details can be found in phase 1 and in^{3,4}). This equation helps to countercheck the number of parameters left in the selected model which is free from multicollinearity and from insignificant variable. It presented a equation to

calculate the number of parameters and demonstrated their application on models with and without interaction variables.

As can be seen from previous section, it requires lengthy time to calculate the number of parameters in a model, especially for bigger models like Eq. 9-10. Instead of calculating the number of parameters one by one manually, the equation established in this study allow researchers to obtain the number of parameters is an easier, faster yet accurate way. Besides, human errors that are caused by manual counting, (have happened during model development) can also be minimised and avoided. The proposed equation also helps to save tremendous amount of time, where there are analysis involving complex iterations or repeated tasks, especially in software development.

REFERENCES

1. Ramanathan, R., 2002. *Introductory Econometrics with Applications*. 5th Edn., Harcourt College Publishers, Ohio, USA., ISBN-13: 9780030341861, Pages: 688.
2. Gujarati, D.N. and D.C. Porter, 2009. *Basic Econometrics*. 5th Edn., McGraw-Hill Companies Inc., New York, USA., ISBN-13: 978-0073375779, Pages: 944.
3. Zainodin, H.J., A. Noraini and S.J. Yap, 2011. An alternative multicollinearity approach in solving multiple regression problem. *Trends Applied Sci. Res.*, 6: 1241-1255.
4. Yahaya, A.H., N. Abdullah and H.J. Zainodin, 2012. Multiple regression models up to first-order interaction on hydrochemistry properties. *Asian J. Math. Stat.*, 5: 121-131.
5. Anastasopoulos, P.C. and F.L. Mannering, 2009. A note on modeling vehicle accident frequencies with random-parameters count models. *Accid. Anal. Prev.*, 41: 153-159.
6. Mazzocchetti, D., A.M. Berti, G. Lucchetti, R. Sartini, F. Colle, P. Iudicone and L. Pierelli, 2012. Evaluation of volume and total nucleated cell count as cord blood selection parameters. *Am. J. Clin. Pathol.*, 138: 308-309.
7. Jaccard, J., 2001. *Interaction Effects in Logistic Regression*. SAGE Publications, London, Pages: 70.
8. Lind, D.A., W.G. Marchal and S.A. Wathen, 2012. *Statistical Techniques in Business and Economics*. 15th Edn., McGraw-Hill, New York.
9. Chittawatanarat, K., S. Pruenglampoo, V. Trakulhoon, W. Ungpinitpong and J. Patumanond, 2012. Development of gender-and age group-specific equations for estimating body weight from anthropometric measurement in Thai adults. *Int. J. Gen. Med.*, 5: 65-80.
10. Zeileis, A., 2004. Econometric computing with HC and HAC covariance matrix estimators. *J. Statist. Software*, 11: 1-17.
11. Ansoorge, H.L., S. Adams, A.F. Jawad, D.E. Birk and L.J. Soslowsky, 2012. Mechanical property changes during neonatal development and healing using a multiple regression model. *J. Biomechan.*, 45: 1288-1292.
12. Emuchay, C.I., S.O. Okeniyi and J.O. Okeniyi, 2014. Correlation between total lymphocyte count, hemoglobin, hematocrit and CD4 count in HIV patients in Nigeria. *Pak. J. Biol. Sci.*, 174: 57-573.
13. Furnstahl, R.J. and B.D. Serot, 2000. Parameter counting in relativistic mean-field models. *Nucl. Phys. A*, 671: 447-460.
14. Robinson, P.S., T.W. Lin, A.F. Jawad, R.V. Iozzo and L.J. Soslowsky, 2004. Investigating tendon fascicle structure-function relationships in a transgenic-age mouse model using multiple regression models. *Ann. Biomed. Eng.*, 32: 924-931.
15. Festing, M.F., 2006. Design and statistical methods in studies using animal models of development. *ILAR J.*, 47: 5-14.