Mathematical Theory and Modeling
ISSN 2224-5804 (Paper) ISSN 2225-0522 (Online)
Vol.8, No.4, 2018
www.iiste.org

Effects of Some Coding Techniques On Multicolinearity and Model Statistics

[1*]Eze Francis C, [2]Nwankwo Chike H, [3]Nwosu Lazarus O, [4]Igweze Amechi. H

1,2,3,4 Department of Statistics, Nnamdi-Azikiwe University P.M.B 5025 Awka, Anambra State, Nigeria

[*]E-mail: ezefcc@yahoo.com

Abstract

Two known methods of coding data for analyses in the presence of multicollinearity and evaluation of model performance viz: Dummy coding and Effect coding which are alternatives to each other were considered. Efforts were made to improve on their performances by modifying them as modified Dummy coding and modified Effect coding respectively and their performances of the now coding methods compared in this paper. The results show that all coding methods significantly reduced the effect of multicollinearity. The effect coding was found to be the best coding method in remedying multicollinearity while closely followed by the dummy coding. However, the proposed modified dummy coding gave the best R-squared values as well as F-values while still reducing the effect of multicollinearity to a great extent and closely followed by modified effect coding. The dummy and effect coding methods proved very efficient in remedying multicollinearity as their observed variance inflation factor (VIF) were all close to unity.
Keywords: Dummy coding, effect coding, multicollinearity, variance inflation factor.

1. Introduction

Coding methods refer to ways in which membership in a group can be represented in mutually exclusive and exhaustive manner. In general, any categorical variable with k categories can be represented by creating (k-1) dummy variables that take on numerical values. This process involves assigning one numerical value, which is called a code, to all subjects of a particular group and different numerical value to all those of the other groups, this is because data need to be represented quantitatively for the purpose of analyses and since categorical variables lack this property (Keppel &Zedeck, 1989).

Grotenhuis, et al (2017) posits that there are many coding methods available and popular is 'dummy coding' in which the estimates represent deviations from a preselected reference category. A way to avoid choosing a reference category is effect coding, where the resulting estimates are deviations from a grand (unweighted) mean. An alternative for effect coding was given by Sweeney and Ulveling in 1972, which provides estimates representing deviations from the sample mean and is especially useful when the data are unbalanced (i.e., categories holding different numbers of observation).

Several statistical coding methods are abound in literature. Generally, any categorical variable can be represented numerically by coding group or category membership as 0's and 1's. Any variable coded in this manner is a dummy coded vector. When only one dummy coded vector is used in a regression equation, the overall regression results indicate whether there is a relationship between the dummy vector and the criterion variable **Y**. When dummy variables are used in regression, the interpretation of the regression coefficients is as follows: the intercept, β_0, represents the mean of **Y** for those coded 0 on the dummy vector, and β_1 represents the mean difference between the group coded 1 and the mean of all those coded 0 on the vector **X**.

According to O'Grady and Medoff (1988), dummy coding yields the same sum of squares as other coding techniques but only under some specific circumstances. These are
(a) if the analysis does not involve any interaction terms
(b) if the analysis of orthogonal design in which tests of significance is tested with a method where the variance is associated with a given predictor and adjusted for the effects in some specific subset of the other prediction variables in the equation, and
(c) analysis of non-orthogonal designs in which the variance is associated in the same way as in (b).

Effect coding was developed out of the desire to test all category means against one overall mean value (Hardy 1993). By doing so one avoids preselecting a reference category as in dummy coding. Effect coding is well-suited whenever the data are balanced, i.e., when the numbers per category of a nominal or ordinal variable are (roughly) equal. Effect coding is very similar to dummy coding and is created with the same process as in dummy coding with the exception that the last group will be assigned −1 for all contrasts, so only k−1 contrasts will be used.

Sundström (2010) discussed coding schemes and coding techniques, noting that one useful aspect with coding schemes is that qualitative data can be changed into quantitative data to make mathematical calculations possible. Another issue is that large amounts of data that normally would take a lot of time to calculate can be transformed into 1's and 0's to make the calculations more effective. To be able to create a coding scheme it is important to have a clear vision of what questions are to be answered. If the researcher does not have a clear vision of what he wants to investigate it might be difficult to choose the coding technique that is best suited in that specific case. If the researcher does not have a clear vision of the problem the result might also be difficult to interpret.

The purpose of the research conducted by Karim (2013) was to determine whether the use of different data coding give different results in the estimation of consumer choice model. The results of the analysis indicate that both dummy and effect coding produce similar results in terms of the model goodness of fit and coefficient of price. However, the estimated coefficients are different. The estimation model that used dummy coding seems to produce better results based on the total number of significant coefficients. Hence calculation using dummy coding was more reliable. Based on the results, the use of dummy coding is preferred in the case where the estimation model does not include intercept. The finding suggests that the interpretation of estimates using different coding should be done with caution as it gives different results which can leads to different policy implications.

Starkweather (2010) addressed the importance of choosing a reference category in dummy coding. The control group represents a lack of treatment and therefore is easily identifiable as the reference category. The reference category should have some clear distinction; however, much research is done without a control group. In those instances, identification of the reference category is generally arbitrary, but Garson (2006) offers some guidelines for choosing the reference category. First, using categories such as miscellaneous or other is not recommended because of the lack of specificity in those types of categorizations. Second, the reference category should not be a category with few cases, for obvious reasons related to sample size and error. Thirdly, some researchers choose to use a middle category, because they believe it represents the best choice for comparison rather than comparisons against the extremes.

Starkweather (2010) demonstrated four strategies for coding a categorical predictor variable for inclusion in linear regression. Each offers specific utility for researchers implementing quasi-experimental designs and true experimental designs. The study noted that each of these strategies resulted in identical values for model summary statistics, and argued that this would not be the case if multiple predictor variables were included in the model. Each of these strategies is compatible with multiple predictors, either continuous or categorical, which highlights the importance of understanding the differences associated with each strategy. The interpretation of regression coefficients differs across each strategy. The study also gave a cautionary note about the use of categorical variables in regression. Given the preceding comment about the predicted values being the same across strategies, it should be clear that regression works best with continuous rather than categorical variables. However, if multiple predictors are included in the model, the use of categorical predictors becomes more precise. Because, instead of predicting the mean of each category (which was represented here due to only having one predictor), the predicted values resulting from the model will be based on all the variables included in the model.

Alkharusi (2012) described how categorical independent variables can be incorporated into regression by virtue of two coding methods: dummy and effect coding. The paper discussed the uses, interpretations, and underlying assumptions of each method. Their findings reveal that the overall results of the regression are unaffected by the methods used for coding the categorical independent variables. The analysis tests whether group membership is related to the dependent variables. Both methods yield identical R^2 and F values. However, the interpretations of the intercept and regression coefficients depend on what coding method has been applied and whether the groups have equal sample sizes.

2. Methodology

The data used in this study are data on Federal Government Recurrent Expenditure and the Gross Domestic Product of Nigeria. Recurrent expenditure refers to expenditure, which does not result in the creation or acquisition of fixed assets. It consists mainly of expenditure on wages, salaries and supplements, purchases of goods and services and consumption of fixed capital while the Gross Domestic Product (GDP) is the monetary value of goods and services produced in an economy during a period of time irrespective of the nationality of the people who produced the goods and services. It is calculated without making deductions for depreciation. The recurrent expenditure comprises of Administrative expenses, Social and Community expenses, expenses on Economic Services and transfers.

Table 1. Recurrent Expenditure of Federal Government of Nigeria and GDP

Year	Administration (X_1)	Social and Community Services (X_2)	Economic Services (X_3)	Transfers (X_4)	GDP(Y)
1981	0.91	0.29	0.18	3.46	144.83
1982	1.04	0.33	0.20	3.93	154.98
1983	0.90	0.29	0.17	3.39	163.00
1984	1.10	0.35	0.21	4.16	170.38
1985	1.43	0.46	0.27	5.41	192.27
1986	1.45	0.47	0.28	5.50	202.44
1987	3.84	0.30	0.69	10.81	249.44
1988	5.78	2.11	1.22	10.30	320.33
1989	6.27	4.23	1.42	14.07	419.20
1990	6.54	3.40	1.61	24.67	499.68
1991	6.95	2.68	1.30	27.31	596.04
1992	8.68	1.34	3.08	39.93	909.80
1993	30.57	14.66	7.75	83.75	1,259.07
1994	20.54	10.09	3.91	55.44	1,762.81
1995	28.76	13.82	5.92	79.13	2,895.20
1996	46.55	15.99	4.75	57.20	3,779.13
1997	56.18	22.06	6.20	74.12	4,111.64
1998	50.68	21.44	11.57	94.40	4,588.99
1999	183.64	71.37	87.08	107.58	5,307.36
2000	144.53	84.79	28.59	203.69	6,897.48
2001	180.80	79.63	53.01	265.86	8,134.14
2002	266.51	152.19	52.95	225.15	11,332.25
2003	307.97	102.61	96.07	477.65	13,301.56
2004	306.77	134.39	58.78	610.70	17,321.30
2005	434.67	151.65	64.31	670.60	22,269.98
2006	522.20	194.17	79.69	594.05	28,662.47
2007	626.36	256.67	179.07	527.17	32,995.38
2008	731.02	332.93	313.75	739.66	39,157.88
2009	714.42	354.19	423.61	635.75	44,285.56
2010	1,117.44	550.90	562.75	878.34	54,612.26
2011	1,262.40	785.44	310.50	956.18	62,980.40

2012	1,159.40	790.06	230.10	1,145.60	71,713.94
2013	1,111.82	844.07	291.23	967.83	80,092.56
2014	992.84	774.77	266.40	1,392.93	89,043.62
2015	1,228.99	807.62	275.36	1,520.01	94,144.96

NBS Annual Bulletin

2.1 Testing for Multicollinearity

The presence of multicollinearity was tested using the Variance Inflation factor (VIF) technique. Wonsuk et al (2014) defined variance inflation factor as a measure of how much the variance of the estimated regression coefficient b_i is "inflated" by the existence of correlation among the predictor variables in the model. According to the author, a VIF of unity means that there is no correlation among the ith predictor variable and the remaining k-1 predictor variables, hence the variance of b_i is not inflated at all. The general rule is that VIFs exceeding 10 are signs of serious multicollinearity requiring correction. The VIF will be used to test the presence of multicollinearity in the data and also used to measure the effect of the various coding techniques on multicollinearity.

The variance inflation factor for a specific variable X_i is given by:

$$VIF_i = \frac{1}{1-R_i^2} \qquad i = 1,...,k \tag{1}$$

Where R^2 is the (coefficient of determination) value obtained by regressing the ith predictor on the remaining predictors**.**

2.2 Dummy Coding

The dummy coding method is given as:

$$X_{it}^{(d)} = \begin{cases} 1, \text{ if } X_{it} > X_{i(t-1)} \\ 0, \text{ otherwise} \end{cases} \tag{2}$$

Where;

X_{it} is the ith variable at time t

$\mathrm{i} = 1, 2, 3, 4;\ \mathrm{t} = 1981, 1982, ..., 2015$

and

$X_{it}^{(d)}$ is the dummy code (ie 0 or 1) for the ith variable at $t;\ \mathrm{i} = 1, 2, 3, 4;\ \mathrm{t} = 1981, 1982, ..., 2015$

Table: 2 Result of Dummy Coding

Year	X_1	X_2	X_3	X_4
1981	0	0	0	0
1982	1	1	1	1
1983	0	0	0	0
1984	1	1	1	1
1985	1	1	1	1
1986	1	1	1	1
1987	1	0	1	1
1988	1	1	1	0
1989	1	1	1	1
1990	1	0	1	1
1991	1	0	0	1
1992	0	0	1	1
1993	1	1	1	1
1994	0	0	0	0
1995	1	1	1	1
1996	1	1	0	0
1997	1	1	1	1
1998	0	0	1	1
1999	1	1	1	1
2000	0	1	0	1
2001	1	0	1	1
2002	1	1	0	0
2003	1	0	1	1
2004	0	1	0	1
2005	1	1	1	1
2006	1	1	1	0
2007	1	1	1	0
2008	1	1	1	1
2009	0	1	1	1
2010	1	1	1	1
2011	1	1	1	1
2012	0	1	0	1
2013	0	1	1	0
2014	0	0	0	1
2015	1	1	1	1

2.3 Effect coding

The effect coding method is given as:

$$X_{it}^{(e)} = \begin{cases} 1, \text{ if } X_{it} > X_{i(t-1)} \\ -1, \text{ otherwise} \end{cases} \quad (3)$$

Where

X_{it} is as defined in (2)

$X_{it}^{(e)}$ is the effect coded variable of the ith variable at time t.

Table3. Result of Effect Coding of the Data

Year	(X_1)	(X_2)	(X_3)	(X_4)
1981	0.00	0.00	0.00	0.00
1982	1.00	1.00	1.00	1.00
1983	-1.00	1.00	-1.00	-1.00
1984	1.00	1.00	1.00	1.00
1985	1.00	1.00	1.00	1.00
1986	1.00	1.00	1.00	1.00
1987	1.00	-1.00	1.00	1.00
1988	1.00	1.00	1.00	-1.00
1989	1.00	1.00	1.00	1.00
1990	1.00	-1.00	1.00	1.00
1991	1.00	-1.00	-1.00	1.00
1992	1.00	-1.00	1.00	1.00
1993	1.00	1.00	1.00	1.00
1994	-1.00	-1.00	-1.00	-1.00
1995	1.00	1.00	1.00	1.00
1996	1.00	1.00	-1.00	-1.00
1997	1.00	1.00	1.00	1.00
1998	-1.00	-1.00	1.00	1.00
1999	1.00	1.00	1.00	1.00
2000	-1.00	1.00	-1.00	1.00
2001	1.00	-1.00	1.00	1.00
2002	1.00	1.00	-1.00	-1.00
2003	1.00	-1.00	1.00	1.00
2004	-1.00	1.00	-1.00	1.00
2005	1.00	1.00	1.00	1.00
2006	1.00	1.00	1.00	-1.00
2007	1.00	1.00	1.00	-1.00
2008	1.00	1.00	1.00	1.00
2009	-1.00	1.00	1.00	-1.00
2010	1.00	1.00	1.00	1.00
2011	1.00	1.00	1.00	1.00
2012	-1.00	1.00	-1.00	1.00
2013	-1.00	1.00	1.00	-1.00
2014	-1.00	-1.00	-1.00	1.00
2015	1.00	1.00	1.00	1.00

2.4 Modified-dummy coding

The modified-dummy coding is a modified dummy coding method that incorporates the dummy coding technique as deviations of variables from the mean values. In this we code 1 if the observed value is greater than the mean of the variable and "0" otherwise.

The modified-dummy coding method is given as:

$$X_{it}^{(md)} = \begin{cases} 1, \text{ if } X_{it} > \bar{X}_t \\ 0, \text{ otherwise} \end{cases} \quad (4)$$

Where;

X_{it} is the ith variable at time t; $\text{i} = 1, 2, 3, 4$; $\text{t} = 1981, 1982, ..., 2015$

and

$X_{it}^{(md)}$ is the modified-dummy coded variable of the ith variable at time t.

$\bar{X}_t$ is the mean of X_t

2.5 Modified-effect coding

The modified-effect coding is a modified effect coding method that incorporates the effect coding technique as deviations of variables from the mean value. In this we code 1 if the observed value is greater than the mean of the variable and "-1" otherwise.

The modified-effect coding method is given as:

$$X_{it}^{(me)} = \begin{cases} 1, \text{ if } X_{it} > \bar{X}_t \\ -1, \text{ otherwise} \end{cases} \quad (5)$$

Where

$X_{it}^{(me)}$ is the modified-effect coded variable of the ith variable at time t.

$\bar{X}_i$ is the mean of X_i.

Table 4. Modified Dummy Coding of the Data

Year	X_1	X_2	X_3	X_4
1981	0	0	0	0
1982	0	0	0	0
1983	0	0	0	0
1984	0	0	0	0
1985	0	0	0	0
1986	0	0	0	0
1987	0	0	0	0
1988	0	0	0	0
1989	0	0	0	0
1990	0	0	0	0
1991	0	0	0	0
1992	0	0	0	0
1993	0	0	0	0
1994	0	0	0	0
1995	0	0	0	0
1996	0	0	0	0
1997	0	0	0	0
1998	0	0	0	0
1999	0	0	0	0
2000	0	0	0	0
2001	0	0	0	0
2002	0	0	0	0
2003	0	0	0	0
2004	0	0	0	0
2005	1	0	0	0
2006	1	1	0	0
2007	1	1	1	0
2008	1	1	1	0
2009	1	1	1	1
2010	1	1	1	1
2011	1	1	1	0
2012	1	1	1	0
2013	1	1	1	0
2014	1	1	1	0
2015	1	1	1	0

Table 5. Modified - Effect Coding of the Data

Year	X_1	X_2	X_3	X_4
1981	-1	-1	-1	-1
1982	-1	-1	-1	-1
1983	-1	-1	-1	-1
1984	-1	-1	-1	-1
1985	-1	-1	-1	-1
1986	-1	-1	-1	-1
1987	-1	-1	-1	-1
1988	-1	-1	-1	-1
1989	-1	-1	-1	-1
1990	-1	-1	-1	-1
1991	-1	-1	-1	-1
1992	-1	-1	-1	-1
1993	-1	-1	-1	-1
1994	-1	-1	-1	-1
1995	-1	-1	-1	-1
1996	-1	-1	-1	-1
1997	-1	-1	-1	-1
1998	-1	-1	-1	-1
1999	-1	-1	-1	-1
2000	-1	-1	-1	-1
2001	-1	-1	-1	-1
2002	-1	-1	-1	-1
2003	-1	-1	-1	1
2004	-1	-1	-1	1
2005	1	-1	-1	1
2006	1	1	-1	1
2007	1	1	1	1
2008	1	1	1	1
2009	1	1	1	1
2010	1	1	1	1
2011	1	1	1	1
2012	1	1	1	1
2013	1	1	1	1
2014	1	1	1	1
2015	1	1	1	1

2.6 Coefficient of Determination (R^2)

The coefficient of determination (denoted by R^2) is a key output of regression analysis. It is interpreted as the proportion of the variance in the dependent variable that is predictable from the independent variables. With linear regression, the coefficient of determination is also equal to the square of the correlation between **X** and **Y** scores. R^2 value of zero "0" means that the dependent variable cannot be predicted from the independent

variable. On the other hand, an R^2 of 1 means the dependent variable can be predicted without error from the independent variable. The value of R^2 ranges between 0 and 1 inclusive. The formula for computing the coefficient of determination for a linear regression model with one independent variable is given as:

$$R^2 = \frac{\text{explained variation}}{\text{total variation}} = \frac{SSR}{SST} \qquad 0 \le R^2 \le 1 \tag{6}$$

2.7 Percentage Change in VIF

The VIF Change is a measure of percentage change in the VIF of coded variables as against the VIF of the actual variables. The percentage change is computed as the VIF of the actual variable less VIF of Coded variable divided by the VIF of the actual data multiplied by 100. That is:

$$VIF_{\Delta} = \frac{(Actual\ VIF) - (Coded\ VIF)}{Actual\ VIF} X100 \tag{7}$$

3. Results of Analyses

3.1 Effects of the coding methods on multicollinearity.

Table 6. Coding Methods and Resulting Multicollinearity

	VIF values and Percentage Change in VIF for each X_i				
Coding Methods	X_1 (VIF_{Δ})	X_2 (VIF_{Δ})	X_3 (VIF_{Δ})	X_4 (VIF_{Δ})	**Mean VIF** ($\boldsymbol{mean}\ VIF_{\Delta}$)
Actual	52.913	22.66	6.787	12.489	**23.71**
Dummy	1.472 (-97.22)	1.144 (-95)	1.553 (-77.12)	1.147 (-90.82)	1.329 (-94.40)
Effect	1.475 (-97.21)	1.077 (-95.28)	1.475 (-78.27)	1.157 (-90.74)	1.296 (-94.53)
Modified effect	7.85 (-85.15)	14.29 (-36.9)	7.641 (+12.5)	1.21 (-90.3)	7.75(-67.31)
Modified dummy	11.314 (-78.62)	14.286 (-36.9)	7.429 (+9.46)	4.46 (-64.3)	9.37(-60.48)
BEST	Dummy	Effect	Effect	Dummy	**Effect**

From Table 6, the results reveal that both the dummy and effect coding methods have approximately the same effect on multicollinearity. The dummy and effect coding reduced VIF (multicollinearity) by at least 77% and as high as 97%.

The modified-effect and modified-dummy reduced VIF (multicollinearity) by a reasonable proportion but not compared to the dummy and effect coding methods. The modified-dummy reduced VIF (multicollinearity) by at least 36.9% up to about 78.%. Similarly, the modified-effect also reduced multicollinearity by at least 36.9% up

to about 90.3%. There are however instances (in X_3) when the modified-dummy coding and modified- effect coding increased VIF value by about 9 and 12% respectively.
A comparison of the mean VIF results and mean VIF change shows that the effect coding method gave the best result on remedying multicolinearity as it reduced VIF of the actual data by an average of 94.53%. This was closely followed by the dummy coding method which reduced VIF by an average of 94.40%. The proposed modified dummy and modified effect coding methods also reduced the effect of multicolinearity by average of 60.5% for modified dummy and 67.3% for modified effect coding respectively.
The effect coding was found to be the best coding method among the competing methods studied.

3.2 Effects of the methods on model statistics

Five multiple regression analysis using the design matrix obtained from the raw data as well as those obtained from the four coding methods each were determined. The model statistics indicating level of model significance (F-value) and extent of model fit (R - square) as well as the p-values of F were inspected (see table 7 below).

Table7: Model Statistics

DATA FORM	R-SQUARE	F-VALUE	P-VALUE (of F)
Actual Data	0.989	692.184	0.00
Dummy Coding	0.154	1.369	0.268
Effect Coding	0.150	1.323	0.284
Modified-Dummy	0.859	45.567	0.00
Modified- Effect	0.852	43.181	0.00

A comparison of the coefficients of determination (R - Square) for the various models show that among the four coding methods, the modified dummy gave the highest R - Square value of 0.859 followed by the modified effect with R - Square value of 0.852 respectively. These high R-Square values show that the model fits, using the modified dummy coding and the use of the modified effect coding, engendered better fits to the data than the dummy and the effect coded models.
On the other hand, the F-values of the regression models involving the various coding methods were compared. The results show that the modified-dummy coding method gave the highest F-statistics of 45.567 with a p-value of 0.00 followed by modified-effect with F-value of 43.181 and p-value of 0.00. The effect coding gave the least F-value of 1.323 and p-value of 0.284 followed by dummy coding with F-value of 1.369 and p-value of 0.268.
These results go to corroborate the significantly larger R^2 values for both the modified dummy coding and the modified effect coding methods. These large F-values indicate that a large proportion of the regression relationship between the GDP and the recurrent expenditure was extracted by these modified coding methods.

4 DISCUSSION OF FINDINGS

The results of the analysis have been presented in the previous sections. On the effect of coding methods on multicollinearity, the effect coding techniques was adjudged to be the best followed by the dummy coding. Also on the effect of coding on model statistics, the two proposed coding methods gave the highest R^2 values, highest F-statistics values and least p-values, these are attributes of a good model fit.
A comparison of the results of the proposed coding methods against the existing ones (Dummy and effect coding) may suggest that the proposed methods may not have fully reduced the effect of multicollinearity. In this, the highly inflated variances, even though reduced, have not been reduced significantly. This is the reason for having high R^2 values as well as high F-statistics values together with the significant p-values. Again the high rate of reduction in the VIF of the variables by the dummy coding and effect coding methods suggest that the effect of multicollinearity may have been fully removed thus resulting in the reduced model statistics that were originally inflated by the presence of multicollinearity.
In the case for expenditure on economic services, X_3, where the proposed methods (modified effect coding and modified dummy coding) slightly increased the VIFs instead of reducing it. This suggest that there is a limit to which the proposed methods can reduce the effect of multicollinearity and start increasing again. It may also be related to the nature and distribution of the transformed data.

REFERENCES

[1] Alkharusi, H (2012). Categorical Variables in Regression Analysis: A Comparison of Dummy and Effect Coding. International Journal of Education, Vol. 4 (2)
[2] Garson, G. D. (2006). *Statnotes: Topics in multivariate analysis: Multiple regression.* Retrieved June 3, 2010, from North Carolina State University, Department of Public Administration Web site: http://www2.chass.ncsu.edu/garson/pa765/regress.htm

[3] Grotenhuis, M. Pelzer, B. Eisinga, R. Nieuwenhuis, R. Schmidt-Catran, A.and Konig, R. (2017). When SizeMatters: Advantages ofWeighted Effect Coding in Observational Studies. International Journal ofPublic Health, 62(1):263–167.

[4] Hardy MA. Regression with dummy variables. Newbury Park: Sage; 1993.
Hasan-Basri, B and Karim, M. Z. A (2013). The Effects of Coding on the Analysis of Consumer Choices of Public Parks. World Applied Sciences Journal 22 (4):

[5] Keppel, G., &Zedeck, S. (1989). Data analysis for research designs: Analysis of variance and multiple regression/correlation approaches. New York: W.H. Freeman and Company.

[6] O'Grady,K. E. and Medoff, D. R. (1988). Categorical Variables in MultipleRegression: Some Cautions. Multivariate Behavioral Research, 23

[7] Sundström, S (2010). Coding in Multiple Regression Analysis: A Review of Popular Coding Techniques. U.U.D.M. Project: 14

[8] Starkweather, J (2010). Categorical Variables in Regression: Implementation and Interpretation. Interpretation and Implementation

[9] Sweeney, R. E. and Ulveling, E. F. (1972). A Transformation for Simplifying the Interpretation of Coefficients of Binary Variables in Regression Analysis. The American Statistician, 1972.

[10] Wonsuk, Y, Mayberry, R., Bae, S., Singh, K., Qinghua H., &Lillard, J.W. (2014). A Study of Effects of MultiCollinearity in the Multivariable Analysis. Int J ApplSci Technol. 4(5): 9–19.