Journal of Physical Science and Innovation
ISSN: 2277-0119
Volume 9, No. 4, 2017

APPLICATION OF MULTIPLE REGRESSION MODELS TO ANTHROPOMETRIC DATA

Bartholomew A. Uchendu
Department of Maths/Statistics,
Federal Polytechnic, Nekede, Owerri, nigeria
Email: uchendubartholomew@yahoo.com

INTRODUCTION

In the assessment of the health and nutritional status of communities, body weight is found to be a good index and it is composed of skeleton, muscle and fat. Best set of anthropometric variables reflective of skeleton, muscle and fat that can be utilized in a meaningful way in place of body weight could be traced This is possible through the so-called variable selection regression models, viz step- up, step –down and step- wise. Earlier studies have indicated that in case of pre-school children, step –up and step – down regression models have yielded similar results. However, comparative studies on all the three regression models are rare and there is a need to detect the agreement of the result for the three models.

DATA UTILIZED

The anthropometric data utilized for this purpose consists of height, weight, head, chest, arm and haemoglobin. All the variables have been standardized by taking them as parentages of the international standards.

MATERIALS AND METHODS

Multiple linear regression model is an ideal multivariate statistical model to study the relationship between a response variable and number of explanatory or predictor variables by the following model

$$Y = X\beta + e \qquad 1$$

Where Y is a n x 1 column vector of response variable, X is n x p matrix of p predictor variables of n observations and e is a random error of n x1 dimension associated with Y vector.

The unknown vector estimated, as a solution of the normal equations defined by

$$\hat{\beta} = (X'X)^{-1} X'Y \qquad 2$$

TESTING OF HYPOTHESIS

To examine that, $\beta_1 = \beta_2 = \ldots \beta_p = \beta$, the variance –covariance matrix as given by Draper and Smith (1981).

$$V(\hat{\beta}) = (X'X)^{-1} \sigma^2 \qquad 3$$

Is utilized and an unbiased estimate of variance, σ^2 is calculated by

$$\hat{\sigma}^2 = \frac{YY' - \beta' YY'}{n-p-1} \qquad 4$$

With (n-p-1) degrees of freedom

TEST FOR MULTIPLE CORRELATION COEFFICIENT

To test hypothesis that the multiple regression equation is a good one, the multiple correlation coefficient R or coefficient of determination R^2 is computed as follows.

$$R^2 = \frac{\beta' X'Y - n\bar{Y}2}{Y'Y - nY2} \qquad 5$$

and it is tested with the variance ratio test, where

$$\frac{R^2}{1-R^2} \; \frac{n-p-1}{p} \sim F_{p,\, n-p-1} \qquad 6$$

is F distribution statistic with p, n-p-1 as degrees of freedom? If this statistics is higher than the F value at 5 per cent level or 1 per cent level for (p,n-p-1) degrees of freedom, value of R^2 is taken as significant and the model is a best fit one. Value of R ranges from 0 to 1. Higher the value and closer to the ideal value of'1', better is the model, we have.

STANDARDIZED AND UNSTANDARDIZED REGRESSION COEFFICIENTS

By using equation (1), unstandardized regression coefficients can be derived,. However, by eliminating $\beta_{0,}$ the standardized regression equation can also be worked out. The latter is very useful in knowing the relative importance of each of the predictor variables, the higher the magnitude of the coefficient, it has, and the greater is its contribution. These coefficients will follow normal distribution with zero mean and unit variance.

METHOD FOR SELECTION OF PREDICTOR VARIABLES

The following methods are generally useful (Draper and Smith, 1981; Visweswara Rao et al 1979) for the selection of best set of predictor variables.

BACKWARD ELIMINATION (STEP –DOWN MEHOD) PROCEDURE

The backward elimination method which is also known as step- down method, is more economical than all possible regression methods as it tries to examine only the "best" regressions containing certain number of variables. The basic steps for this method are as follows;

i. A regression equation containing all the variables is computed.

ii. The partial F- test value is calculated for every predictor variable.

iii. The lowest partial F- value, say, $F_{1,}$ is compared with a preselected significance level $F_{0,}$ say,

a. If $F_L < F_{0,}$ the variable X_L is removed, and the regression equation is recomputed with remaining variables

b. If $F_L > F_{0,}$ further calculations are stopped and the latest regression equation is taken as the final equation.

FORWARD SELECTION (STEP- UP METHOD) PROCEDURE

This procedure which is also known as step- up method, involves more computational effort than the backward elimination procedure. In the first step, the explanatory variable, which has the highest and significant simple correlation with the response variable is selected. Then the partial correlation coefficients of the response variable with other variables after including the first variable are calculated, and the variable that has the highest and significant partial correlation coefficient is selected. Again, the partial correlation coefficient of other variables are computed and the variable with highest and significant partial correlation coefficient is selected and the procedure is repeated. The computations are continued as long as the R^2 value increases significantly or the F value associated with the new variable to be introduced is greater than a specified value with consideration.

STEPWISE REGRESSION PROCEDURE

The forward selection procedure has one drawback. A variable which is regarded as the best variable at an early stage may be superfluous at a later stage. In stepwise method, variables are entered into the model using forward selection procedure. At every stage, after selecting a variable, the F values for the variables already selected are calculated and any variable that is superfluous is deleted as explained in step-down regression model. Thus, at each stage, the forward selection procedure is used to decide which variable to include and backward elimination procedure is used to decide which variable to include and backward elimination procedure is used to decide which variable to eliminate.

In order to compare the agreement between the three types of models, the following regression equations have been fitted using the three types of regression procedures, viz step- up, step- down and stepwise for both boys and girls separately and the result are presented in tables 1 to 6.

Response variable	predictor variables
Weight (%)	Height (%), weight for height (%), head (%) Chest (%), arm (%) and haemoglobin (%)

From table 1, it is observed that in the step- up regression model, when weight (%) was regressed with other anthropometric variables for boys, in the first step, height (%) was included in the model with R^2= 0.535, F = 302.36 for (1,263) degrees of freedom. In the second step, the variable weight for height (%) and in third step, head (%) were included though the contribution of the third variables head (%) is negligible. Thus, the final regression model is obtained in terms of just three variables, viz height (%), weight for height (%) and head (%).

In case of step –down regression model for boys (table2), when weight (%) was regressed with other anthropometric variables, in the first step all the variables were included in the analysis. The R^2 value of the model is quite high 0.979 and F value quite significant 2015.53 for (6,258) degrees of freedom.

In step 2, the variable chest (%) was eliminated, the difference in R^2 value with the previous value is nil and there is an increase in F value. In the next two steps, variables hemoglobin (%) and arm (%) were eliminated and the final model was obtained in terms of height (%) , weight for height (%) and head (%). Both the standardized and unstandardized regression coefficient, R^2 value, F value, etc were found to be exactly the same as those obtained in step- up model

From table 3, it is observed that in the case of stepwise regression model for boys, when weight (%) was regressed with other anthropometric variables, height (%) was picked up in the first- step and R^2 value was only 0.535. In the second step, weight for height (%) was also included and a good model could be obtained with R^2 = 0.978 and F= 5954.45 for (2,262) degrees of freedom ($p<0.01$). One more variable head (%) was included in the third step increasing the R^2 value slightly to 0.979 and decreasing the F value to 4015.85 for (3,261) degrees of freedom. Since the increase in the R^2 value is marginal, the equation at step 2 is good enough to all practical purpose. This solution was found to be exactly same as the solution obtained from step- up and step- down models.

In fact, it is observed that in this example, all the three models, viz step – wise, step- up and step- down gave the same parameter estimates, R^2 values and F values.

In the case of girls also all the three models, viz step – up, step- down and stepwise models gave the similar results as given in tables 4 to 6. Thus, all the three models have identified height (%)

Table 1: results of step – up model using weight (%) as response variable and other anthropometric variables as predictor variables – boys (1-5yrs)

Response Variable	Predictor Variable	Step No	B	Beta	R^2	F	Df
Weight (%)	Height (%)	1	1.182	0.731	0.535	302.36*	1,263
	Const		-34.138	-			
Weight (%)	Height (%)	2	1.417	0.878	0.978	5954.45*	2,262
	Wfht (%)		0.894	0.682			
	Const.		-134.058	-			
Weight (%)	Height (%)	3	1.429	0.884	0.979	4015.85*	3,261
	Wfht (%)		0.899	0.686			
	Head(%)		-0.054	-0.019			
	Const.		-130.022	-			

** p< 0.01 Wfht=weight for height (%);*

Head (%) = head circumference (%)

B= Unstandardized regression coefficient, Beta= standardized regression coefficient.

Table 2: Results of step- down model using weight (%) as response variable and other anthropometric variables as predictor variables – boys (1-5yrs).

Response Variable	Predictor Variable	Step No	B	Beta	R^2	F	Df
Weight (%)	Height (%)	1	1.439	0.890	0.979	2015.53*	6,258
	Wfht(%)		0.910	0.694			
	Head (%)		-0.038	-0.013			
	Chest(%)		-0.021	-0.008			
	Arm(%)		-0.025	-0.015			
	HB(%)		-0.006	0.011			
	Const		-128.444	-			
Weight (%)	Height (%)	2	1.432	0.886	0.979	2423.23*	4,260
	Wfht (%)		0.906	0.691			
	Head(%)		-0.041	-0.015			
	Arm(%)		-0.026	0.015			
	HB(%)		-0.006	-0.011			
	Const.		-129.111				
Weight (%)	Height (%)	3	1.431	0.885	0.979	3024.47*	4,260
	Wfht (%)		0.903	0.689			
	Head(%)		-0.046	-0.016			
	Arm(%)		-0.024	-0.016			
	Const.		-129.032				
Weight (%)	Height(%)	4	1.429	0.884	0.979	4015.85*	3,261
	Wfht(%)		0.899	0.686			
	Head (%)		-0.054	-0.019			
	Const		-130.022				

**p< 0.01; head(%)= Head circumference (%); Arm (%) = Arm circumference (%); HB = Hemoglobin level (%), B = unstandardized regression coefficient, beta= standardized regression coefficient.*

Table 3 results of stepwise model using weigth (%) as response variable and other anthropometric variables as predictor variables –boys (1-5yrs)

Response Variable	Predictor Variable	Step No	B	Beta	R^2	F	Df
Weight (%)	Height (%)	1	1.182	0.731	0.535	302.36*	1,263
	Const		-34.138	-			
Weight (%)	Height (%)	2	1.417	0.878	0.978	5954.45*	2,262
	Wfht (%)		0.894	0.682			
	Const.		-134.058	-			
Weight (%)	Height (%)	3	1.429	0.884	0.979	4015.85*	3,261
	Wfht (%)		0.899	0.686			
	Head(%)		-0.054	-0.019			
	Const.		-130.022	-			

**p< 0.01. B = unstandardized regression coefficient;*
Beta= standardized regression coefficient.

Table 4 results of step- up model using weight (%) as response variable and other anthropometric variables as predictor variables- girls (1-5yrs)

Response Variable	Predictor Variable	Step No	B	Beta	R^2	F	Df
Weight (%)	Height (%)	1	0.962	0.649	0.421	160.09*	1,220
	Const		-15.048	-			
Weight (%)	Height (%)	2	1.315	0.887	0.943	1806.09*	2,219
	Wfht (%)		0.807	0.760			
	Const.		-117.400	-			
Weight (%)	Height (%)	3	1.278	0.862	0.945	1257.36*	3,218
	Wfht (%)		0.789	0.743			
	Head(%)		0.159	0.056			
	Const.		-128.637	-			

*p< 0.01, B = Unstandardized regression coefficient, Beta = Standardized regression coefficient

Based on the above studies the following conclusions are made

i. *When weight (%) was regressed with other anthropometric variables using multiple linear regression model, very good prediction equations could be derived.*

ii. *Among the many predictor variables, weight for height (%) and height (%) were found to be the most important combination of variables, a result similar to that observed by Visweswara Rao and colleagues (1976,1979,1986), in case of poor and well-to-do families. This was found to be true in the case of boys as well as girls.*

Table 5 results of step –down model using weight (%) as response variable and other anthropometric variables as predictor variables – girls (1-5yrs)

Response Variable	Predictor Variable	Step No	B	Beta	R^2	F	Df
Weight (%)	Height (%)	1	1.294	0.873	0.948	555.59*	6,214
	Wfht(%)		0.801	0.754			
	Head(%)		0.189	0.067			
	Chest(%)		-0.063	-0.034			
	Arm(%)		0.011	0.023			
	HB(%)		0.013	0.006			
	Const		– 128.094	–			
Weight (%)	Height (%)	2	1.298	0.875	0.978	650.42**	5,215
	Wfht (%)		0.805	0.758			
	Head(%)		0.190	0.068			
	Chest(%)		-0.059	-0.032			
	HB(%)		0.013	0.039			
	Const.		– 128.156	–			
Weight (%)	Height (%)	3	1.281	0.864	0.947	777.17*`	4,216
	Wfht (%)		0.791	0.745			
	Head(%)		0.150	0.053			
	HB(%)		0.014	0.041			
	Const.		– 127.429	–			
WEIGHT	Height (%)	4	1.278	0.862	0.945	1257.36*	3,218
	Wfht (%)		0.789	0.743			
	Head(%)		0.159	0.056			
	Const		– 128.637	–			

B= Unstandardized regression coefficient, Beta= standardized regression coefficient.

iii. The three regression models step- up, step –down and step – wise have given similar results in boys as well as girls (Agreement between these procedure have been found also good with small as well as large samples (Chapter 42).

Table 6 results of stepwise model using weight (%) as response variable and other anthropometric variables as predictor variables –girls (1-5yrs).

Response Variable	Predictor Variable	Step No	B	Beta	R^2	F	Df
Weight (%)	Height (%)	1	0.962	0.649	0.421	160.09*	1,220
	Const		-15.048	-			
Weight (%)	Height (%)	2	1.315	0.887	0.943	1806.21*	2,219
	Wfht (%)		0.807	0.760			
	Const.		-117.400	-			
Weight (%)	Height (%)	3	1.278	0.862	0.945	1257.36*	3,218
	Wfht (%)		0.789	0.743			
	Head(%)		0.159	0.056			
	Const.		-128.637	-			

**p<0.01. B* = Unstandardized regression coefficient, Beta= standardized regression coefficient.

LG_9 – Leg circumference at age 9 (cm)
ST_9 – A composite measurement of strength at age 9 (high values = stronger).
WT_{18}- Height at age 18
LG_{18}- Leg circumference at age 18
ST_{18} – Strength at age 18
SOMA- Somato type, a seven point scale, a measure of fatness (1 slender, 7= fat), determined using a photograph taken at age 18.
Data for 26 boys and 32 girls are given in table 41.7 respectively (the complete study consists of larger sample sizes and of more variables
For both boys and girls obtain the following model:

$$SOMA=\beta_0+\beta_1+HT_2+\beta_2 WT_2+\beta_3 HT_9+\beta_4 WT_9+\beta_5 ST9+e$$

And try to trace the best set of variables by all the three types of models would yield the same type of results or not.

Table 7: Anthropometric measurements of children adolescents by age and sex

BOYS

ID No	WT_2	HT_2	WT_9	HT_9	LG_9	ST_9	WT_{18}	HT_{18}	LG_{18}	ST_{18}	SOMA
201	13.6	90.2	41.5	31.6	31.6	74.0	110.2	179.0	44.1	226.0	7.0
202	12.7	91.4	31.0	144.3	26.0	73.0	79.4	195.1	36.1	252.0	4.0
203	12.6	86.4	30.1	136.5	26.6	64.0	76.3	183.7	36.9	216.0	6.0
204	14.8	87.6	34.1	135.4	28.2	75.0	74.5	178.7	37.3	220.0	3.0
205	12.7	86.7	24.5	128.9	24.2	63.0	55.7	171.5	31.0	215.0	1.5
206	11.9	88.1	29.8	136.0	26.7	77.0	68.2	172.5	39.1	152.0	3.0
207	11.5	82.2	26.0	128.5	26.5	45.0	78.2	172.5	39.1	152.0	6.0
209	13.2	83.8	30.1	133.2	27.6	70.0	66.5	174.6	37.3	189.0	4.0
210	16.9	91.0	37.9	145.6	29.0	61.0	70.5	190.4	33.9	183.0	3.0
211	12.7	87.4	27.0	132.4	26.0	74.0	57.3	173.8	33.3	193.0	3.0
212	11.4	84.2	25.9	133.7	25.8	68.0	50.3	172.6	31.6	202.0	3.0
213	14.2	88.4	31.1	138.3	27.2	59.0	70.8	185.2	36.6	208.0	4.0
214	17.2	87.7	34.6	134.6	30.6	87.0	73.7	178.4	39.2	227.0	3.0
215	13.7	89.6	34.6	139.0	28.9	71.0	75.2	177.6	36.8	204.0	2.5
216	14.2	91.4	43.1	146.0	32.4	98.0	83.1	183.5	38.0	226.0	4.0
217	15.9	90.0	33.2	133.2	28.5	82.0	74.3	178.1	37.8	233.0	2.5
218	14.3	86.4	30.7	133.2	27.3	73.0	72.2	177.0	36.5	237.0	2.0
219	13.3	90.0	31.6	130.3	27.5	68.0	88.6	172.9	40.4	230.0	7.0
221	13.8	91.4	33.4	144.5	27.0	92.0	75.9	188.4	36.5	250.0	1.0
222	11.3	81.3	29.4	125.4	27.7	70.0	65.6	180.2	35.7	236.0	3.0
223	14.3	90.6	30.2	135.8	26.7	70.0	65.6	180.2	35.4	177.0	4.0
224	13.4	92.2	31.1	139.9	27.2	63.0	66.4	189.0	35.3	186.0	4.0
225	12.2	87.1	27.6	136.8	25.8	73.0	59.0	182.4	33.5	199.0	3.0
226	15.9	91.4	32.3	140.6	27.9	69.0	68.1	185.8	34.2	227.0	1.0
227	11.5	89.7	29.0	138.6	24.6	61.0	67.7	180.7	34.3	164.0	4.0
228	14.2	92.2	31.4	140.0	28.2	74.0	68.5	178.7	37.0	219.0	2.0
GIRLS											
331	12.6	83.8	33.0	13.5	29.0	57.0	71.2	169.6	38.8	107.0	6.0
334	12.0	86.2	34.2	137.0	27.3	44.0	58.2	166.8	34.3	130.0	5.0
335	10.9	85.1	28.1	129.0	27.4	48.0	56.0	157.1	37.8	101.0	5.0
351	12.7	88.6	27.5	139.4	25.7	68.0	64.5	181.1	34.2	149.0	4.0

352	11.3	83.0	23.9	125.6	24.5	22.0	53.0	158.4	32.4	112.0	5.0
353	11.8	88.9	32.2	137.1	28.2	59.0	52.4	165.6	33.8	136.0	4.0
354	15.4	89.7	29.4	133.6	26.6	58.0	56.8	166.7	32.7	118.0	4.5
355	10.9	81.3	22.0	121.4	24.4	44.0	49.2	156.5	33.5	110.0	4.0
356	13.2	88.7	28.8	133.6	26.5	58.0	55.6	168.1	34.1	104.0	4.5
357	14.3	88.3	38.8	134.1	31.1	57.0	77.8	165.3	39.8	138.0	6.5
358	11.1	85.1	36.0	139.4	28.2	64.0	69.6	163.7	38.6	108.0	5.5
359	13.6	91.4	31.3	138.1	27.6	64.0	56.2	173.7	34.2	134.0	3.5
361	13.5	86.1	33.3	138.4	29.4	73.0	64.9	169.2	36.7	141.0	5.0
362	16.3	94.0	36.2	139.5	28.0	52.0	59.3	170.1	32.8	122.0	4.5

ID No	WT_2	HT_2	WT_9	HT_9	LG_9	ST_9	WT_{18}	HT_{18}	LG_{18}	ST_{18}	SOMA
364	10.2	82.2	23.4	129.8	22.6	60.0	49.8	164.2	30.0	128.0	4.0
365	12.6	88.2	33.8	144.8	28.3	107.0	62.6	176.0	35.8	168.0	5.0
366	12.9	87.5	34.5	138.9	30.5	62.0	66.6	170.9	38.8	126.0	5.0
367	13.3	88.6	34.4	140.3	31.2	88.0	65.3	169.2	39.0	142.0	5.0
368	13.4	86.9	38.2	143.8	29.8	78.0	65.9	172.0	35.7	132.0	5.5
369	12.7	86.4	31.7	133.6	27.5	52.0	59.0	163.0	32.7	116.0	5.5
370	12.2	80.9	26.6	123.5	27.2	40.0	47.4	154.5	32.2	112.0	4.0
371	15.4	90.0	34.2	139.9	29.1	71.0	60.4	172.5	35.7	137.0	4.0
372	12.7	94.0	27.7	136.1	26.7	30.0	56.3	175.6	34.0	114.0	3.0
373	13.2	89.7	28.5	135.8	25.5	76.0	61.7	167.2	35.5	122.0	4.5
374	12.4	86.4	30.5	131.9	28.6	59.0	52.4	164.0	34.8	121.0	5.0
376	13.4	86.4	39.0	130.9	29.3	38.0	58.4	161.6	33.0	107.0	6.5
377	10.6	81.8	25.0	126.3	25.0	50.0	52.8	153.6	33.4	140.0	5.0
380	12.7	91.4	29.8	135.5	27.0	57.0	67.4	173.5	34.5	123.0	5.0
382	11.8	88.6	27.0	134.0	26.5	54.0	56.3	166.2	36.2	135.0	4.5
383	13.3	86.4	41.4	138.2	32.5	44.0	82.8	162.8	42.5	125.0	7.0
384	13.2	94.0	41.6	142.0	31.0	56.0	68.1	168.6	38.4	142.0	5.5
385	15.9	89.2	42.4	140.8	32.6	74.0	63.1	169.2	37.9	142.0	5.5

And weight for height (%) as the most important variables. All the three models have picked up head (%) variable also, even though it is not significant. The R^2 value is 0.945 and the associated F value for (3,218) df is 1257. 36, which is highly significant ($p<0.01$) tables 4 to 6

NOTE

ID NO - Identification number

WT_2 - Weight at age 2 (kg)

HT_2 - Height at age 2 (cm)

WT_9 -Weight at age 9

HT_9 - Height at age 9

SUMMARY

Stepwise, step-down and step-up procedure provided similar results in the selection of predictor variable. Height with weight for height (%) was found best for use in place of weight for the nutritional status of children.

REFERENCES

Draper NR, Smith H, (1981) Applied regression Analysis. Second edn. John Wiley and Sons..

Jelliffe D.B.(1966).Assessment of the nutritional status of the community. WHO monograph series; 53:pp.63-74.

Visweswara Rao K, Balakrishna N, Adinarayana K (2006). Critical limits of some anthropometric measurements and indices for the assessment of nutritional status, Indian. J Nutr and Dietet 1986; 23:pp 88-99

Weisberg S. Applied linear Regression. Second edn. John Wiley and sons 1985; pp267-71.

Reference to this paper should be made as follows: Uchendu, Bartholomew A (2017), Application of Multiple Regression Models to Anthropometric Data. J. of Physical Science and Innovation, Vol. 9, No. 4, Pp. 1-17