

Exploring Body Dimensions to Predict Scale Weight of Teenagers in Imo State, South-East, Nigeria.

Anyiam Kizito E. and Uchegbulem-Asiegbu Nwaoma P.

Abstract- In this work, the multiple linear regression models were used to explore the relationship in body dimensions and to predict scale weight for teenagers in Imo state, Nigeria. The scale weight was predicted via anthropometric measurements as well as age, height and gender for 604 physically active secondary school students-283 boys and 321 girls within the age range of 10-21 years (in completed years). The work resulted in twelve different linear regression models. The ordinary regression procedure was used to obtain the first nine models while the forward stepwise selection procedure was employed to check the performance of first three models. Studying the effects of skeletal diameter, girth circumference measurement and gender, on the response variable (scale weight). The dataset was split accordingly and the results of the individual models were compared with the full model. The goodness and adequacy of the fitted models were studied on the basis of F-value, t-value, R-squared criterion, standard error of estimate and Mallows' C-P statistics. Results of the analysis showed that body scale weight is better estimated by anthropometric, skeletal and girth variables as well as age, height and gender than any of the reduced sets of variables. The result also indicated that all the models were of good fit and adequate.

Index Terms: Anthropometric Data, Body scale weight, forward stepwise selection procedure, Model Comparison, Multiple Linear Regression

1 Introduction

Anthropometric data are very important data which researchers use for different professional, academic, social and nutritive purposes. The data in use to represent anthropometric measurements are in many different formats as body dimensions vary significantly according to the population of origin. For gender classification, Lotman et al (1998) and Heinz et al (2003), could classify gender using bicromial diameters. In forensic science, Joyce and Store (1991), Wingate (1992), and Heinz et al (2003) claimed that gender of adults could be determined accurately up to 70% or more using the skeletal remains of their anthropometric measurements mainly pelvic and skull measurement. According to the argument of Heinz et al (2003), White and Churchill (1975), Clauser et al (1977), Centre for Disease Control (CDC) office (Undated but assessed 2004), the anthropometric dataset like the one in this study could be used in ergonomic design equations if completed with data on body segment length. Aroskar and Panas (2004), Abdali et al (2004) through multiple regression analysis explored relationship in body dimensions for 507 physically active adults using anthropometric dataset which is similar the data set of this study. They also used anthropometric dataset as well as age, height and gender to predict scale weights of these individuals. There what predict scale weight best and other reasons are what motivated this study. Twelve regression models for weight were fitted to predict scale weights of teenagers using their anthropometric

measurements as measures of their body dimensions as well as their ages, heights and gender. In doing these, the dataset was categorized into three namely: all variables (all dataset of skeletal and girth measurements) as well as height. i.e. skeletal and girth body dimensions, age and gender; reduced variables (girth measurements i.e. girth body dimensions) as well as height, age and gender; reduced variables (skeletal measurements i. e. skeletal body dimensions) as well as height, age and gender.

The Pearson correlation matrix ex-arrayed the relationship in body dimensions of the teenagers. The models were fitted for boys and girls, and both. The models fitted for both sexes are the focus. Multivariate multiple linear regression method was employed to fit the models. Ordinary regression and stepwise procedures were adopted to access and select the best models.

2.0 Theory of Multivariate multiple Regression Analysis

Let $X_{i1}, X_{i2}, \dots, X_{it}$ be multivariate random variables (such as skeletal (diameter) and girth circumference measurements as well as height, age and gender) assumed to be related to a response variable Y (body weight). The predictor variables of interest X determined by a mean which depends continuously on the values of the $X_{ij}, i = 1, 2, \dots, m, j = 1, 2, \dots, t$ and the error δ_{ij} are

$$\hat{\Sigma} = Y - \hat{Y} = [I - X(X'X)^{-1}X']Y \tag{2.6}$$

The orthogonality conditions among the residuals predicted model (values) and columns of X also hold since,
 $X'[I - X(X'X)^{-1}X'] = X' - X' = 0$; and

$$X'\hat{\Sigma} = X'[I - X(X'X)^{-1}X']Y = 0$$

$$\hat{Y}'\hat{\Sigma} = \hat{\beta}'X[I - X(X'X)^{-1}X']Y = 0 \text{ since,}$$

$$Y = \hat{Y} + \hat{\epsilon} \text{ i.e. } Y'Y = (\hat{Y} + \hat{\Sigma})'(\hat{Y} + \hat{\Sigma}) = \hat{Y}'\hat{Y} + \hat{\Sigma}'\hat{\Sigma},$$

which by implication,

$$\hat{\Sigma}'\hat{\Sigma}' = Y'Y - \hat{\beta}'X'X\hat{\beta}$$

We recall the best linear unbiased property of $\hat{\beta}$, $E(\hat{\beta}) = \beta$ and $Cov(\beta_i, \beta_k) = C_{ik}^2 (X'X)^{-1}$; $i, k = 1, 2, \dots, t+1$,
 with residuals

$$\hat{\Sigma} = [\hat{\Sigma}_{(1)} \quad \hat{\Sigma}_{(2)} \quad \dots \quad \hat{\Sigma}_{(k)}] = Y - X\hat{\beta}$$

satisfying (i) $(\hat{\Sigma}_{ik}) = 0$; (ii) $E(\hat{\Sigma}'\hat{\Sigma}') = \frac{n-t-1}{C_{ik}^2}$ so that

$$E\left(\frac{\hat{\Sigma}'\hat{\Sigma}'}{n-t-1}\right) = C_{ik}^2 I = \Sigma$$

For multivariate multiple linear regression, the hypothesis

$$H_0 : \beta_1 = \beta_2 = \dots = \beta_k = 0$$

$$H_1 : \beta_{(i)} \neq 0, i = 1, 2, \dots, k \text{ for some } i$$

Is investigated using any of the following; F-test, t-test, Hotelling's T²-test, all of which according to Garson (2005) are special cases of Wilks Lambda test and likelihood ratio test expressed in generalized inverse form which is equivalent to Wilks Lambda Statistic (Johnson and Wichem (1992)). For details of t-test and F-test and other procedures for fitting regression models, see Neter et al (1996).

In this study, the Mallows' C_p Statistics was employed to assess the adequacy of the regression models and to select the 'best' regression equation, see Statsoft (2003) and Neter et al (1996). The Mallows C_p-statistic is of the form

3.2 Analysis

$$C_p = \frac{RSS_p}{S^2 - (n - 2p)} \tag{2.7}$$

Where RSS_p is the residual sum of squares from a model containing P parameter including β₀. The S² is the residual mean square from the largest equation postulated containing all the X_(is). It is important to recall that a regression model with how C_p about equal to P is concluded to be free from lack of fit.

3.0 Materials and Methods

Multivariate multiple linear regression analysis was employed to explore body dimensions to predict scale weight of teenagers with a view of selecting the best model

3.1 Data Collection and Description

In the study, locally fabricated anthropometer, plastic tape, scale weight and extended meter rule were used to collect anthropometric measurements (variables) from twenty- one body sites of six hundred and four (604) physically active teenagers within the age range of (10-20) years. The dataset consist of these anthropometric variables measurement in centimeters as well as heights (centimeters), weights (in kilogrammes), ages (in completed years), and gender of these six hundred and four teenagers. The measurement techniques and the body sites used were as suggested and prescribed by Belinke and Wilmore (1994). Nine skeletal (diameter) measurements (variables) included in the dataset were measured using the anthropometer. The skeletal variables include those measured at indicated body sites bicromial (BCRL), biiliac (BILC), bitrochanteric (BTCT), Chest (CHTD), elbow (ELBD), wrist (WRTD), knee (KNED), ankle (ANKD) diameters and chest dept (CHTP). The twelve girth (circumference) measurements (variables) used in this study are the three bony girths of the wrist (WRTGT), knee (KNEGT), ankle (ANKGT) and other nine girths measured at the sites shoulder (SHDGT), Chest (CHTGT), Waist (WATGT), Navel (NAVGT), Hip (HPGT), Thigh (THGT), Bicep (BCPGT), Forearm (FRAGT) and Calf (CAFGT), all measured with plastic tape. Each respondent had his/her age, weight, height, gender and anthropometric measurements recorded given a total of twenty-four predictor variables and one response variable (weight).

The multivariate multiple linear regression analysis was used to analyze the data since the data are multivariate data satisfied the usual assumptions. In the normal and stepwise regression analysis procedures were used. Nine (9) different regression models were run using the normal regression procedures with statistical packages for Social sciences Version 10.0. The stepwise regression procedures were used to run three (3) different models giving a total of twelve (12) regression models. The first three models run using normal regression procedure and the last three models using stepwise selection procedure will be presented in this study. For others, the correlation matrix of the variables, plots of histograms for some of the variables, normal p-p plots of regression statistics for the models, the dataset (ANPE data), results of parameter estimation and contribution of the variables to weight in each of the models (see, Uchebulam, 2006).

4.0 Result and Discussion

The six models discussed are presented in full

Model 1: Full model, all dataset (n=604) all variables (24) Model FMADAV.

The regression equation is:

$$wt = -22.6215 + 2.8868AG - 0.9185HT + 0.8564CHTD + 0.5923CHTP + 0.6540ELBD + 1.7878WRD + 0.421KNED + 0.3075ANKD + 0.3789BCRL + 0.7682BILC + 0.4015BTCT + 0.1845WRTGT - 0.1727KNEGT - 0.0669ANKGT + 0.2696SHDGT + 0.2971CHTGT - 0.0902WATGT + 0.3064NAVGT + 0.1549HPGT + 0.1437THGT - 0.2553BCPGT + 0.4591FRAGT - 0.0624CAFGT \quad (4.1)$$

With $R^2=80.78\%$, R^2 -adjusted = 80.02% , S. E. = 10.015, $C_p=26$, F-test value =0.000, and residual mean square of 100.10. Most influential explanatory variable on the basis of their t- values are age, height, chest diameter, chest depth, elbow diameter, chest girth, forearm, wrist diameter, navel girth, biiliac and bicromial diameters.

Model 2: Full models, all dataset (604), reduced variables, (skeletal variable= 9, age and height)

Model FMADRVS:

The weight function for this model is

$$Wt = 6.5914 + 2.8046AG - 0.8950HT + 1.0856CHTD + 1.0316CHTP + 0.4712ELBD + 1.2082WRD + 1.3722KNED + 0.8543ANKD + 0.6233BCRL + 1.1965BILC + 0.4593BTCT \quad (4.2)$$

WITH $R^2=77.23\%$, R^2 -adjusted = 76.61% , S. E. =10.79, $C_p=12$, F-test value= 0.000 and residual mean square of 116.45. The most contributing predictor variables on the basis of their t-test are age, height, chest diameter, chest depth, elbow diameter, knee diameter, biiliac and bicromial diameters.

Model 3: Full model, all dataset (604), reduced variables (girth variables = 12, age and height)

Model FMADRVG

The weight function for this model is:

$$Wt = 1.0895 + 3.2541AG - 0.8566HT + 1.3276WRTGT + 0.5569KNEGT + 0.3479ANKGT + 0.3931SHDGT + 0.1936CHTGT - 0.0010WATGT + 0.1427NAVGT + 0.2704HPGT + 0.1410THGT + 0.3239BCPGT + 0.6315FRAGT - 0.1103CAFGT \quad (4.3)$$

With $R^2=76.12\%$, R^2 -adjusted = 75.56% , S. E. 11.098, $C_p=15$, P-value=0.000 and residual mean square of 122.7166. The most influential predictor variables are age, height, forearm, hip, chest and knee girth.

The three models are all of good fit judging by their F-value and C_p values. The predictor variables predicted the response variable in each case (model) with sufficient accuracy as indicated by the R^2 and R^2 -adjusted criteria. However, model FMADRVS is a better model than the model FMADRVG on the basis of R^2 and R^2 -adjusted criteria, while model FMADAV is better than both models on the same criteria. This result suggest that separating the predictor variables into skeletal variables and girth variables does not improve the quality of the regression model thereby confirming the hypothesis that scale weight is better predicted by skeletal and girth body dimensions (variables) as well as age and height.

Assessing the performance of these models obtained through ordinary regression procedure was used to run the same models and the results obtained are hereby presented.

Model 4: Full model, all dataset (n=604) all variables (24), using stepwise (spw)

Mdel SPW – FMADAV

Twelve different regression models were obtained and the optimum on the basis of R^2 , R^2 -adjusted and S. E. is

$$Wt = -21129 - 0.9144HT + 2.996AG + 0.483BCRL + 0.216THGT + 0.498FRGT + 1.075BILC + 0.333NAVGT + 0.910CHTP + 0.651CHTP + 0.5342ELBD + 0.309CHGT + 2.037WRTGT \quad (4.4)$$

Having $R^2=80.3\%$, R^2 -adjusted= 79.9% and S.E. 10.053 with the most influential predictor variables on the basis of their t-test values are age, height, forearm girth, bicromial diameter, naval girth, chest diameter, chest depth, elbow diameter, chest girth and wrist diameter. It is important to note that the

variables dropped finally in this model were those rendered redundant by the presence or (inclusion) of some other variables which are highly correlated with them. For instance ankle diameter and wrist diameter was favoured. For the same reason, the thigh girth which is strongly correlated with bicep girth, hip girth and waist girth was favoured

Model 5: Full model, all dataset (n=604), reduced variables skeletal variables (9), age and height, using stepwise regression

Model SPW-FMADRVS

Eight regression functions were obtained and the optimum for the group being

$$Wt = 8.720 - 0.891HT + 2.952AG + 0.632BCRL + 1.457KNED + 1.512BILC + 0.495ELBD + 1.004CHTP + 1.050CHTP \quad (4.5)$$

With R²=76.9%, R²-adjusted=76.6%, S. E. 10.84 and residual mean square of 117.530. out of the nine (9) skeletal variables only six (6)

Table 4.1: Summary of the results

Model	No of variables	R ² (%)	R ² -adjusted (%)	S. E.	C _p	F- value	Pr>F/sign.F
FMADAV (1)	24	80.76	80.02	10.0152	26	106.01	<0.001
FMADRVS (2)	12	77.23	76.81	10.7911	12	182.53	<0.001
FMADRVG (3)	15	76.12	75.56	11.0778	15	134.14	<0.001
Spw-FMADAV (4)	24	80.3	79.9	10.0531	-	200.38	0.000
SPW-FMADRVS (5)	12	76.9	76.6	10.8411	-	247.61	0.000
SPW-FMADRVG (6)	15	76.0	75.5	11.0806	-	187.27	0.000

Comparing these six models which we obtained using two different regression procedures on the whole dataset (n=604), we observed that elimination of the predictor variables considered redundant by the stepwise regression procedure did not improve the quality of the fit. This suggests that the predictor variables not included in the models by this procedure may not be completely uninfluential. Models 2, 3, 5 and 6 obtained with reduced predictor variables did not show any better fit either. Comparing models 1 and 4 the two full models we noticed that model 1 (FMADAV) is a better model. This suggests that the influences of all the explanatory variables are important in predicting body scale weight.

appeared in the optimum regression function. This again is as a result of strong correlation between and among body dimensions. Bitrochanteric diameter was out because of the correlation with biiliac diameter.

Model 6: Full model, all dataset, (n=604), reduced variables (girth) (12), age and height using stepwise.

Model SPW- FMADRVG

Twelve different regression functions were obtained and the basis of the usual criteria the optimum for this group is:

$$Wt = 1.134 - 0.856HT + 3.266AG + 0.621FRAGT + 1.503WRTGT + 0.231NAVGT + 0.415SHDGT + 0.20HPGT + 0.178CHTGT + 0.573KNEGT + 0.339ANKGT \quad (4.6)$$

With R²=76.0%, R²-adjusted=75.5%, S. E.= 11.081 and residual mean square error = 112.780. Only eight girth variables of forearm, wrist, naval shoulder, hip, chest, knee and ankle were most influential.

4.1 Model Comparison

The six regression models obtained in this study using normal and stepwise regression procedures compared well with the models obtained using best subset and forward- backward regression selection procedures by Aroskar et al (2004) whose dataset was similar to the dataset of this study but for their age range of (21-30) years and geographical location. The R² statistic is sufficiently high for all the fitted models in this study as was the case with the models of study conducted by Aroskar et al (2004). This is an indication that the models of this study are good fits for the data, agreeing with the conclusions of Aroskar et al (2004).

Similarly, Heinz et al (2003) identified model 2 of their study as body build equation for weight for their respondents. This

comparison is adequate because the skeletal variables are assumed to be highly consistent over years are not affected by changes in the body fat or muscle mass. Judging from model 2 of

4.2 Conclusion

In this study, we have explored body dimensions of skeletal and girth anthropometric measurements to predict scale weight of teenagers via multiple linear and stepwise regression approach. The analysis showed that all the models are 'good fit' for the data and are adequate judging by their R^2 , R^2 - adjusted values. Also shown by the study is that reduction in the variables to skeletal or girth alone or using stepwise regression selection procedure did not improve the quality and adequacy of the fitted models. Thus body scale weight is better predicted significantly by anthropometric measurements of both skeletal and girth body dimensions as well as age, height and gender than any other group of reduced variables.

Finally, the models of this study compared well with the models of Aroskar et al (2004) in Georgia Atlanta, U. S. A. and Heinz et al (2003) in California, U. S. A.

References

Abdali O., Vikhtor H., Paquet E. and Ritous N. Exploring Anthropometric Data through Cluster Analysis. www.iit-iticnrc.gc.ca/iit-publications.iti/docs/NRC-46564.ppt, (2004).

Aroskar, N. and Panes A Exploring Relationships in Body Dimensions. [Regression Analysis]. Trran Projects for ISYE 644, Statistical Models and Regression Analysis. www.projects.amaleshpanse.com/stat/reporte. Pdf, (2004).

Behnke A. R. and Wilmore J. H. Evaluation and Regulation of Body Build and Composition. Eaglewoods Citiffs, N. J: Prentice Hall, (1994).

Clauser, L; Tucker, P; McConville, J. Churchill, E; Laubach, L; and Reardon, J. Anthropometry of Force Women, Report Number AMRL-TR-70-5. Aerospace Medical Research Laboratory. Wright-Patterson Air Force Base, Ohio, (1972).

Draper, N., R. and Smith H.. Applied Regression Analysis. 3rd ed. Newyork. John Wiley and sons, (1998)

Garson, G. D. GLM: MANOVA and MANCOVA. www2.chess.ncsu.edu/garson/pg765/manova.htm, (2005).

Heinz, G., Peterson L. J. Johnson R. W and Lork C. J.. Exploring Relationships in Body Dimensions. Journal of Statistics Education: Vol. 11. No. 2, (2003) www.Amstat.org/publications/jse/v11n2/datasets.heinz.htm/

Johnson, R. A. and Wichern, D. W. Applied Multivariate Statistical Analysis. 3rd ed. Prentice Hall. Inc. New Jesey. (1992)

Heinz et al (2003) and model FMADRV5 of this study it was observed that the girth measurements in model 2 of Heinz et al (2003) are those not affected by body fat and adult age.

Joyce, C., and Stove E Witness from the grave. The stories Bones Tell. Boston. MA. Little, Brown and Company Pg.177-178, (1991).

Lohman, T. Roche, A. and Morderell, R (1988). Anthropometric Standardization Reference Manuel. Campaign II: Human Kinetics Books. Multiple Regression, Stat. soft, inc, (1994-2003).

Neter, J.; Kutner, M. N.; Nachitsterm, G, J. and Wassermen, W. Applied Linear Statistical Models. MCB/MC. Graw- Hill a division of the Mc: Graw-Hill Companies, U. S. A., (1996)

Nwobi, F. N. and Nduka, E. C Statistical Notes and Tables for Research. 2nd ed. Afro-Ortis Publications Nsukka, Nigeria, (2003).

White, R. N. and Churchill, E. The Body Size of Soldiers: U S Army. Anthrometry - 1966 Report Number T2-51-CE (CPLSEL-94). U. S Army Natick Laboratories/ Natick, M. A., (1971).

Wingate, A. Science of the Crime: A writer's guide to Crime-science investigations, Cincinate, OH; Writer's Digest Books, P. 148, (1992).

- Author name is currently pursuing masters degree program in electric power engineering in University, Country, PH-01123456789. E-mail: author_name@mail.com
- Co-Author name is currently pursuing masters degree program in electric power engineering in University, Country, PH-01123456789. E-mail: author_name@mail.com
(This information is optional; change it according to your need.)