Artificial Neural Network: A New Approach for Prediction of Body Fat Percentage Using Anthropometry Data in Adult Females

Sugandha Mehandiratta[1] Priyanka Singhal[2] A.K. Shukla[3] and Rita Singh Raghuvanshi[4]
[1]Ex-PG Student, Department of Mathematics, Statistics and Computer Science, College of Basic Sciences and Humanities
[2]Ex-PG Student, Department of Foods & Nutrition, College of Home Science,
[3]Dean, College of College of Basic Sciences and Humanities
[4]Dean, College of College of Home Science
(G. B. Pant University of Agriculture & Technology, Pantnagar, Uttarakhand, 263145, India.)

Abstract: Assessing body fat using anthropometric data would be useful in predicting chronic diseases. Accurate use of proper statistical models in analysing body composition data is of prime importance. This study was undertaken to assess body composition of diseased and non-diseased women using body composition analyser thereafter using data for development of statistical model. The objective was to find relationship of various anthropometric parameters with Percent Body Fat (BF%) and to develop various prediction models for estimating BF on the basis of anthropometric data. BF% was predicted using Linear Regression (LR), Multiple Linear Regression (MLR), Non-Linear Regression (NLR) and Artificial Neural Network (ANN) models. The predictors used in the study were age (yrs.), height (cm), weight (kg), Body Mass Index (BMI) (Kg/m^2) and Waist Circumference (WC) (cm). Data utilized for the study was related to 860 adult females aged 18-60 years out of which 700 were non-diseased and 160 were diseased (diabetic and hypertensive). Out of various models developed using LR, MLR, NLR for Non-Diseased group, three predictors viz. age, BMI and WC were found to be appropriate for estimating BF%. However, the best prediction of BF% was achieved using ANN model taking age, height, weight and WC as predictors (R^2 = 0.787). ANN technique was found as the most suitable technique for developing prediction models for estimation of BF% in non-diseased group. However, in diseased group ANN model could not predict BF% more precisely, may be due to some other factors affecting the body composition of females of diseased group.

Key Words: *Body fat: Anthropometry: Artificial Neural Network: Women.*

*****

I. Introduction:

Anthropometry is an important aspect used for understanding physical variation and assessing body composition. Despite the modern techniques, anthropometric measurements are important to study the genetic structure and prediction of risk factors of many complex diseases in human health. Anthropometric parameters such as height, weight, body mass index, and waist circumference evaluate underweight and obesity conditions. Anthropometric measurements are easy to conduct in field studies and have better acceptance in the communities.

Overall obesity and visceral obesity in which a high proportion of body fat is deposited on the trunk and in the abdomen are associated with deleterious health outcome. The prevalence of obesity is increasing rapidly worldwide. The causes of obesity are not completely understood but according to heredity and decreased physical activity are its major causes (1). It is observed that reduction in lean body mass and relative increase of percentage body fat mass is due to the age related change in chemical composition of an individual (2). Age is associated with regional fat distribution causes negative health consequences. Body fat gets redistributed with age. Visceral obesity is predicted outcome of increased waist circumference. Waist circumference rises with advancing age. It has been reported that among older group larger waist circumference than that of younger one (3).

An increased risk of developing cardiovascular disease, chronic disorders and disabilities was determined in the case of females with a WC over 80 cm, and a strongly increased risk in the case of females with a WC of over 88 cm was observed (4). Mean value for fat (%) was observed as 19.0±4.4 and 29.2±5.6 cm at WC cut off points viz., <80 cm and ≥80 cm respectively. It shows presence of higher body fat in abdominally obese female adults (5). The use of statistical models in medical diagnoses and biomedical research may affect individual's life by predicting whether their health is protected or is in danger. Therefore, careful and accurate use of proper statistical models in analysing health related data is of prime importance.

BF% is the percent of total body weight of a person. BF % can be predicted from BMI (6). The ideal weight and BF% varies for males and females of different age groups. Excess body fat contributes to an array of medical diseases and can greatly increase the risk of contracting conditions such as coronary heart disease, diabetes, cancers, gallstones etc. Deurenberg et al developed prediction equations based on anthropometric parameters for both males and females using LR technique. It was also found that same models could not be used for both males and females because BF% varies with gender [7]. Jackson et al developed equations for both males and females separately considering only one variable as predictor i.e. BMI [8]. In studies carried out by scientists[9-10] it was observed that BF% also depends on other anthropometric parameters too rather than just BMI. Several studies were also made by various scientists [11-13] for estimating BF% using anthropometric data.

In the present study, an attempt has been made to develop prediction models for estimating BF% using ANN which is a new and advanced. The developed models were also compared with the models developed by classical statistical techniques like LR, MLR, NLR to measure the precision in the model. Assessment of nutritional status is an important is an essential component for development of public health programme and policies. With increase in obesity and degenerative diseases, body composition is being considered for more sensitive methods for predicting degenerative disease. This paper will help in assessing body fat using anthropometric data which is accepted in communities.

II. Subjects and Methods:

A cross-sectional study was conducted on 860 adult females aged between 18-60 years recruited by personal contact. Out of 860 adult females; 700 were physically fit (non-diseased) and 160 were diabetic and hypertensive (diseased). Selection of non-diseased female participants was made from University premise using simple random sampling without replacement technique. Self reported diabetes is estimated to be over 2 per cent among women as per national family health survey. Prevalence of hypertension has been found as 10 per cent among rural population as per Registrar General of India [14]. Therefore, national prevalence estimates for hypertension and diabetes were taken into account to determine the sample size for the present investigation. The approximate diseased group required to estimate prevalence at 95 per cent level of confidence with a precision level of 5 per cent was 138. Keeping in view increased prevalence with advancing years, therefore, 160 women population constituted the sample size. Diseased subjects were selected from University hospital using purposive sampling. According to NFHS-3 (2005-2006) approximately 36 per cent of women are undernourished signifying high nutritional deficiency. Prevalence of under-nutrition and over-nutrition was taken into consideration for calculation of sample size for non-diseased population. Approximately 13 per cent of women are overweight or obese. Adults in India are suffering from a dual burden of malnutrition[16]. Therefore, a table value of 637 was taken, considering that 99 per cent confidence interval with a precision level of 5 per cent will give expected prevalence of 40 per cent (under-nutrition).Both diseased and non-diseased populations were considered to have sedentary life style. This study was conducted according to the guidelines laid down in the Declaration of Helsinki and all procedures involving human subjects were approved by the committee.

Anthropometric measures such as height, weight and WC were recorded using standard techniques. Body fat was assessed through bioelectrical impedance analysis using bioscan analyser (Maltron, 916). For body composition assessment, study subjects were asked to lie down in a supine position. ME4000 electrode pads were placed clipped with MEC1104 electrode cable on a hand and foot and consequently, a small alternating electrical current (50 kHz) is passed through these electrodes. The impedance to the current flow is measured. The electrical impedance of body tissues is measured by BIA and used to have estimates on fluid volumes, total body water, body cell mass and fat free mass [17]. Diseased subjects were diagnosed through conducting biochemical test using standard methods and procedures. Blood pressure values of 120/88 mm hg for systolic and diastolic pressure was taken as normal value while beyond that indicated as high blood pressure. Blood pressure was measured by automatic blood pressure meter. A blood glucose level of 120 milli grams per decilitre as less after an 8 hour fast was considered as fasting blood glucose whereas the higher values were considered as high blood sugar level. Study subjects were clearly explained about purpose of research investigation. All the subjects gave their free and written informed consent to be part of the study. All the procedures used were non-invasive in nature. Separate statistical analysis was carried out for non-diseased and diseased group.

III. Statistical Analysis:

Comparison of various anthropometric parameters between non-diseased and diseased group was done using two sample Z-test, descriptive statistical measures like mean and standard error of mean ($S.E_m$) were also calculated. Karl Pearson's correlation coefficient and partial correlation coefficient were used for determining relationships between various body composition and anthropometric parameters. Significance of correlation coefficient and partial correlation coefficient was tested using t-test. LR, MLR, NLR and ANN techniques were used to develop various prediction models based on age and other anthropometric parameters. Coefficient of multiple determination (R^2) was used for the comparison of developed models. F-test was used to test the significance of Coefficient of Multiple Correlation (R).

IV. Results:

a) Two sample Z-test and descriptive statistical measures: The values of Zcal, mean and $S.E_m$ of age, height, weight, BMI and WC for non-diseased and diseased group are presented in **Table 1**. It can be observed that average height, weight, BMI and WC differ significantly between non-diseased and diseased group and were found to be higher on an average in diseased group as compared to non-diseased group. This observation signifies that age and diseased state affect anthropometric indices. Larger mean values were obtained in persons with diabetes, hypertension and in subjects above 30 years of age**.**

b) Karl Pearson's correlation coefficient: Correlation analysis results for non-diseased and diseased group are shown in **Table 2**. Age and anthropometric measurements including weight, BMI and WC were significantly positively correlated with BF% for both the groups (non-diseased and diseased). However, height was significantly negatively correlated with BF% in case of Non-Diseased group.

There were high positive and significant correlations between weight and BF% (r=0.732) & BMI and BF% (r=0.782) in case of non-diseased group as compared to diseased group. Height was not significantly correlated with BF% in case of diseased group.

c) Partial correlation coefficient: In order to find the net association between BF% and anthropometric parameters an attempt was made to find partial correlations. Partial correlation coefficients between BF% and BMI controlling the effect of weight and age separately as well as together for both non-diseased and diseased group are presented in **Table-3.**

Positive and significant correlation exists between BF% and BMI eliminating the effect of weight and age separately as well as together. It indicates that there is a net association between BF% and BMI irrespective of the effect of age, weight or both the variables on BF% and BMI. However, non-significant partial correlation was observed between BF% and BMI (r=0.197) for diseased group when the effect of weight and age together was controlled.

d) Prediction Models for the estimation of BF% using LR, MLR and NLR: BF% was found to be significantly correlated with age, weight, BMI and WC except height for diseased group. Therefore an attempt has been made to develop prediction equations for estimation of BF% on the basis of explanatory variables (age, height, weight, BMI and WC) using LR, MLR and NLR analysis and these models are presented in **Table 4**.

Five prediction models were developed for non diseased group using LR analysis for estimating BF% on the basis of explanatory variables age, height, weight, BMI and WC (taking one parameter at a time) but for diseased group only four prediction models were developed on the basis of explanatory variables age, weight, BMI and WC (taking one parameter at a time) as in this case it was observed that height was insignificantly correlated with BF%. In LR models the value of R^2 was found to be low ranging from 0.007 to 0.632. Best prediction of BF% could be achieved taking BMI as explanatory variable for which maximum value of R^2 is 0.632 and 0.154 in case of non-diseased and diseased group respectively.

An attempt has been made to develop some models as NLR taking some nonlinear relations like cubic, quadratic, logarithmic, inverse, exponential, power, compound etc. in order to increase the R^2 value. Out of the 10 NLR best fit models developed for estimating BF%, 5 for each group using age, height, weight, BMI and WC taking one variable at a time the best prediction on the basis of R^2 can be achieved by taking only BMI as explanatory variable with cubic relationship between BF% and BMI for both the groups shown in **Table 4**.

Further, an attempt has been made to predict BF% by using different combinations of age, height, weight, BMI and WC as explanatory variables using MLR models. On the basis of R^2 values, the best MLR model developed for the non-diseased group with maximum value of r^2= 0.659 and the best fit MLR model for diseased group was with maximum value of R^2 = 0.184 (**Table 4**).

e) Prediction models for estimating BF% with ANN: A new mathematical technique i.e. (ANN) was also applied to develop prediction models for BF% on the basis of age and anthropometric variables. The proposed neural network was trained using different data sets to get the best set of data. Several ANN models were developed for estimating BF% taking different combinations of explanatory variables age, height, weight, BMI and WC for both the groups(non-diseased and diseased). For each model R^2 value was calculated and its significance was tested.

The construction of the ANN model was divided into following three subparts

i) **Division of the data set:** The data was divided into three parts for each non-diseased and diseased group which is shown in **Table 5**

ii) **Block Structure:** Block Structures used for developing ANN models are shown in **Figure 1** and **Figure 2** for non-diseased and diseased group respectively**.** During the training phase, the system adjusts its connection/weights strengths in favor of the input variables that are more effective in determining a specific output. The use of verification set in study is an important guard against overtraining/over-fitting the network. Two hidden layers and seven/six processing units in hidden layers for non-diseased/ diseased group were tested to find out the right combination of processing units in order to solve the problem with acceptable training times and with optimum precision.

iii) **Scatter plot:** Scatter plot for best developed ANN models indicating predicted and observed values for non-diseased and diseased group for estimating BF% are presented in **Figure 3** and **Figure 4**.

Hyperbolic tangential type ANN model was found to be most suitable for estimation of BF% using four explanatory variables namely age, height, weight, WC with the maximum value of R^2 = 0.787 for non-diseased group (**Figure 1**). However, in case of diseased group Hyperbolic tangential type ANN model was found to be most suitable for estimation of BF% using only three explanatory variables namely age, BMI, WC with the maximum value of R^2= 0.114 (**Figure 4**).

V. Discussion:

Objective of the present investigation was to find the relationship of various anthropometric parameters with BF% and to develop prediction models for predicting BF% on the basis of anthropometric parameters age, height, weight, BMI and WC separately for non-diseased and diseased group of Indian adult females. BF% was positively significantly correlated with all anthropometric measurements for both the groups except height, with which it was negatively significantly correlated for non-diseased group and insignificantly correlated for diseased group. Positive and significant partial correlation was observed between BF% and BMI eliminating the effect of weight and age separately as well as together for non-diseased group. Non-significant partial correlation was observed between BF% and BMI for diseased group when the effects of weight and age together were controlled. Therefore, it can be concluded that there is no association between BF% and BMI for diseased group.

Positive and significant correlation was observed between BF% and BMI for diseased group which may be due to the effect of weight and age on both the variables or due to the effect of some other variables.

LR models based results suggested that best prediction of BF% could be achieved by taking BMI as explanatory variable. Also, BMI was found to be better predictor for estimating BF% for Non-diseased group as compared to diseased group.

Results of MLR models suggest that taking three explanatory variables at a time better prediction of BF% could be achieved. It was concluded that best prediction of BF% could be achieved using age, BMI and WC as explanatory variables for both non-diseased and diseased group. Results of NLR models for prediction of BF% suggests that the best prediction could be achieved by taking only BMI as explanatory variable with cubic (polynomial of third degree) relationship between BF% and BMI for estimating BF%.

Results of ANN models suggests that out of the various ANN models developed, two best fit ANN models were obtained for estimation of BF% incorporating four and three explanatory variables respectively for Non-diseased and diseased group. Hyperbolic tangential type ANN model was most suitable for estimation of BF% using age, height, weight and WC for non-diseased group with maximum value of R^2=0.787. However, for diseased group hyperbolic tangential type ANN Model was most suitable for estimation of BF% using age, BMI and WC with maximum value of R^2= 0.114. Very Low value of R^2 indicates that there are some other factors in the human body influencing the anthropometry data in case of diseased group.

Finally on the basis R^2 values various developed prediction models viz., LR, MLR, NLR and ANN for estimation of BF% were compared and it was concluded that the best prediction of BF% could be achieved by ANN technique followed by MLR technique with three explanatory variables. The best prediction could be achieved by ANN model for non-diseased group as compared to diseased group. Therefore, it is concluded that ANN is a powerful technique for estimating BF% and can be effectively used for further research in the analysis of body composition data. Use of anthropometric data, statistical tool and software may help in predicting normal health and warning signal for onset of disease in the communities without using sophisticated instruments.

Tables-1

Table 1. Subject characteristics

Parameters	Non-Diseased Group (n=700)	Diseased Group (n=160)	
	($\bar{x} \pm$ S.E$_m$)	($\bar{x} \pm$ S.E$_m$)	Zcal
age (yrs)	24.15±7.60	49.22±5.87	46.14**
height (cm)	156.36±6.24	159.39±7.22	6.47**
weight (kg)	50.55±9.26	61.17±11.54	10.851**
BMI(kg/m^2)	20.69±3.67	26.56±4.29	15.57**
WC (cm)	77.84±11.60	97.23±11.10	19.360**
BodyFat(%)	22.85±7.78	39.84±11.95	22.228**

S.E$_m$; Standard Error of Mean,** Significant at α=0.01.

Table-2

Table 2. Karl Pearson's coefficients of correlation of anthropometric parameters with BF%.

Parameters	Correlation Coefficients with BF%	
	Non-Diseased group (n=700)	Diseased group (n=160)
Age	0.365**	0.178*
Height	-0.082*	-0.015
Weight	0.732**	0.216**
BMI	0.782**	0.276**
WC	0.670**	0.170*

* Significant at α=0.05; ** Significant at α=0.01.

Table 3

Table 3. Partial Correlation coefficients between BF % and BMI

Controlled variables	Partial correlation coefficients	
	Non-diseased group (n=700)	Diseased group (n=160)
Age	r = 0.443**	r = 0.219**
Weight	r = 0.768**	r = 0.355**
weight and age	r =0.420**	r = 0.197

** Significant at α=0.01

 ISSN: 2321-8169
117 - 125

Table-4

Table 4: Best Prediction Models for the estimation of BF% using LR, MLR and NLR Models

Technique	Prediction Models: Non-Diseased group (n=700)	Prediction Models: Diseased group (n=160)
LR	BF%= -12.033**+1.686**BMI (R^2=0.632**)	Y=10.774+1.112**BMI (R^2=0.154**)
NLR	**BF%=23.781+2.019BMI+0.029BMI^2-0.001BMI^3 (R^2=0.790**)**	**Y=168.940+17.903BMI+0.741BMI^2-0.009BMI^3 (R^2=0.177**)**
MLR	BF%=-18.273**-0.050age+1.278**BMI+0.172**WC (R^2=0.659**)	Y=109.834**-0.238*age-1.195**BMI-0.079WC (R^2=0.184**)

Table-5

Table 5. Division of the data set for ANN

Model Details		**(Percent)** Non diseased group (700)	**(Percent)** Diseased group (160)
Sample	Training	493 (70.8%)	116 (72.%)
	Testing	203 (29.2%)	44 (28%)
Valid		696	160
Excluded		4	0
Total		700	160

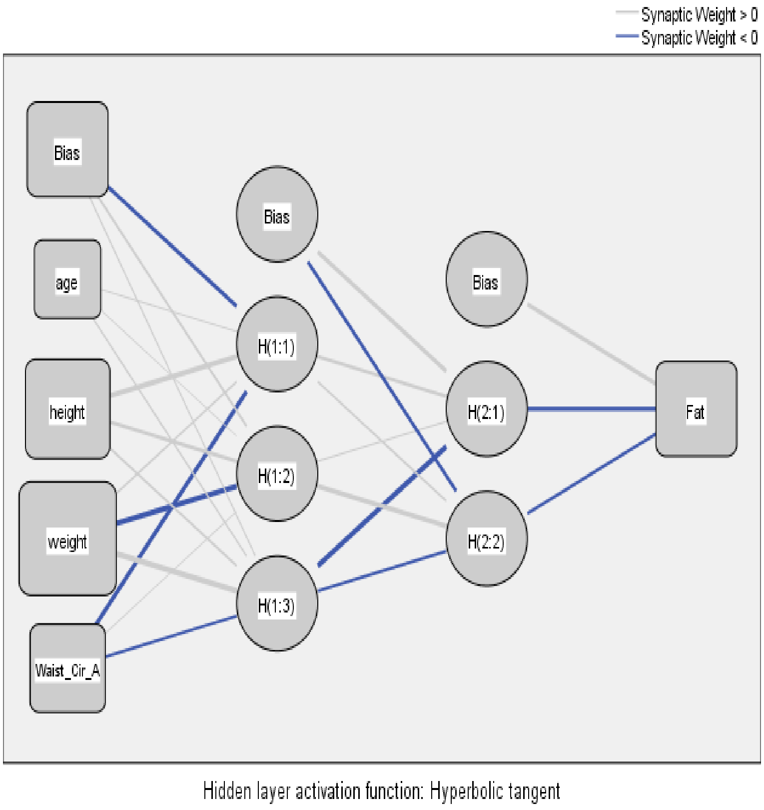

Figure1. Block Structure of ANN Model for non-diseased group

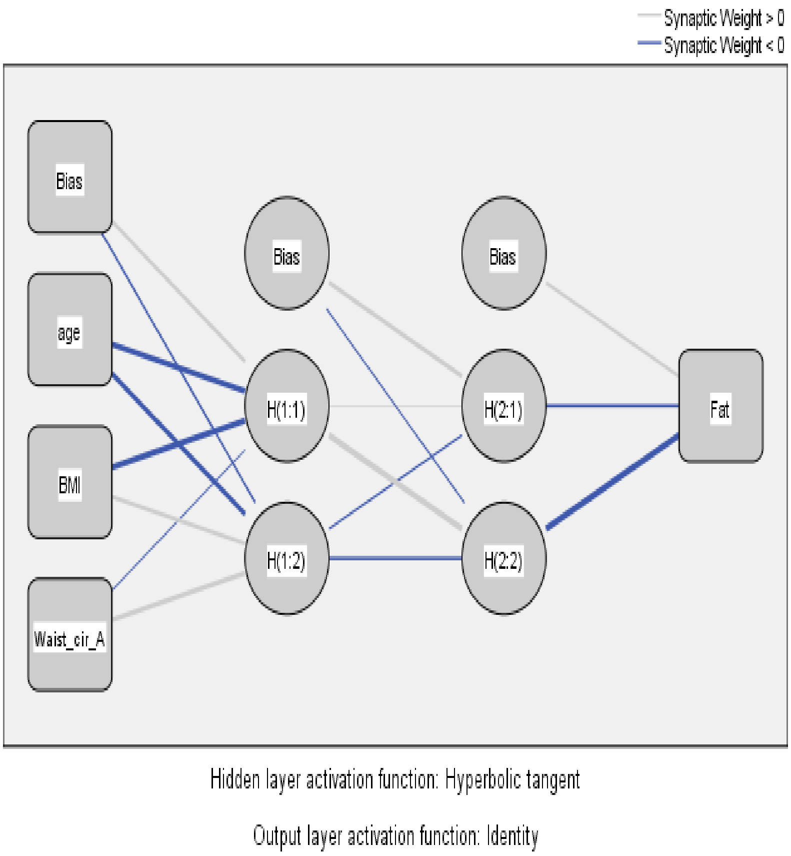

Figure2. Block Structure of ANN Model for diseased group

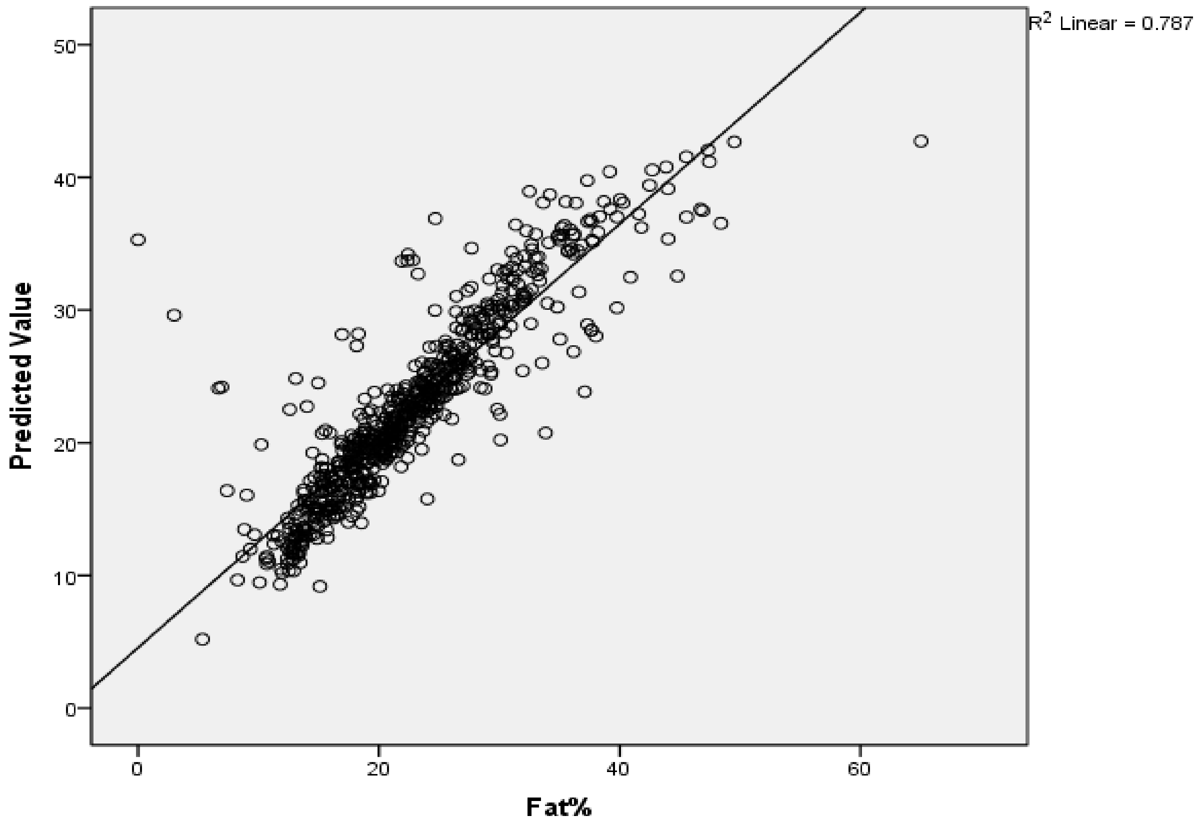

Figure 3. Predicted VS Observed plot for non-diseased group

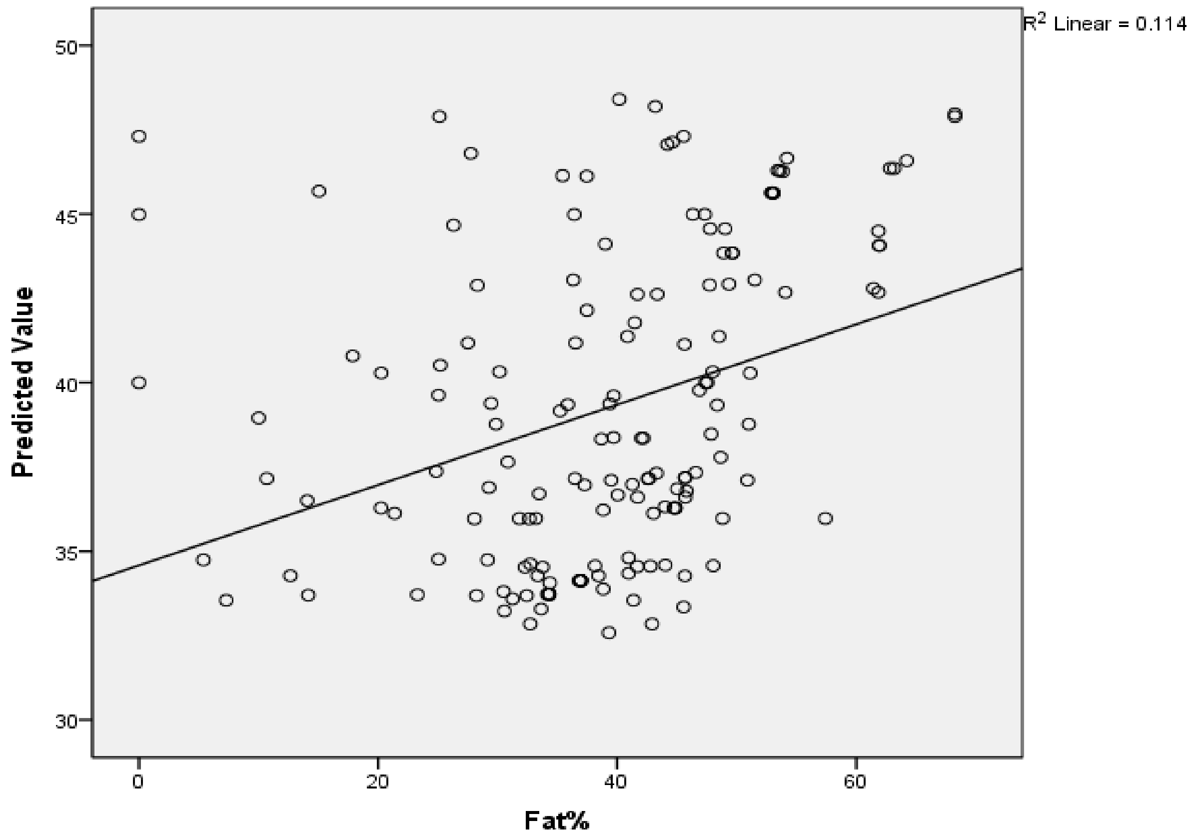

Figure 4. Predicted VS Observed plot for diseased group

ISSN: 2321-8169
117 - 125

References:

[1]. W.H.O (World Health Organization) (1995). The use of interpretation of Anthropometry-Report of a WHO Expert Committee. Geneva 8-11.

[2]. Bemben MG, Massey GH, Bemben DA and Minser JE (1998). Age-related variability in body composition methods for assessment of percent fat and fat-free mass in men aged 20-74 years. *Age and ageing.* 27: 147-153.

[3]. Ford ES, Mokdad AH and Giles WH (2003). Trends in waist circumference among U.S. adults. *Obesity Research.* **11** (10):1223-1231.

[4]. W.H.O (World Health Organization). (1997).Waist Circumference and Waist–Hip Ratio: Report of a WHO Expert Consultation. Geneva, 10–14.

[5]. Raghuvanshi RS and Singhal P (2013). Prevalence of Abdominal Obesity and Its Association With Body Composition Indices In Young Adult Indian Females. *International Journal Food and Nutritional Sciences*; **2** 55-61.

[6]. Norgan NG. (1994). Population differences in body composition in relation to BMI. *Eur J Clin Nutr.* 48: S10–S27.

[7]. Deurenberg P, Deurenberg M and Wang J. (2000). Prediction of percentage body fat from anthropometry and bioelectrical impedance in Singaporean and Beijing Chinese.*Asia Pacific J ClinNutr*.**2**: 93–98.

[8]. Jackson AS, Stanforth PR and Gagnon J. (2002). The effect of sex, age and race on estimating percentage body fat from body mass index: The Heritage Family Study. *International Journal of Obesity*.**26**: 789–796.

[9]. Miyatake N, Miyachi M, Tabata I, Sakano N and Hirao T. (2012). Relationship between muscle strength and anthropometric, body composition parameters in Japanese adolescents.*Health*.4:1-5. (19)

[10]. Sorenson HT, Sabroe S and Sorenson TI. (1999). Birthweight and Length as Predictors for Adult height. *American Journal of Epidemiology*.**8**:726-729.

[11]. Cameron N, Griffiths, PL, Wright MM, Blencowe C, Davis NC, Pettifor JM and Norris SA. *(*2004). Regression equations to estimate percentage body fat in African pre-pubertal children aged 9 y[1-3]. *American journal of Clinical Nutrition*.80: 70 –75.

[12]. Sekulic D, Zenic N and Zubcevic NG. (2007). Non Linear Anthropometric Predictors in Swimming. *Coll. Antropol.* **3**: 803–809.

[13]. Snyman H, Herselmen MG and Labadories D. (2008). The development of a preliminary regression equation for estimating the weight of black South African paraplegic males using anthropometric measurements in Tshwane. *South African Journal of clinical nutrition.***21**: 127-131.

[14]. Report on causes of deaths in India 2001-2003. Office of the Registrar General of India, Govt.of India, 2010.

[15]. Thrusfield, A. (1995). *Veterinary Epidemiology*. II ed. Blackwell science Ltd.

[16]. Arnold F, Parasuraman S, Arokiasamy P and Kothari M. (2009). Nutrition in India. National Family Health Survey (NFHS-3), India, 2005-06. Mumbai International Institute for Population Sciences; Calverton, Maryland, USA: ICF Macro.

[17]. Lukaski, H.C. (1987). Methods for the assessment of body composition: traditional and new. *Am J Clin Nutr*. **46**: 437–456.