FIXING MULTICOLLINEARITY INSTABILITY IN THE PREDICTION OF BODY WEIGHT FROM MORPHOMETRIC TRAITS OF WHITE FULANI COWS

Abdulmojeed YAKUBU

Department of Animal Science, Faculty of Agriculture, Nasarawa State University, Keffi, Shabu-Lafia Campus, P.M.B. 135, Lafia, Nigeria.

Corresponding author's e-mail address: abdul_mojeedy@yahoo.com, Tel. No: +2348065644748

ABSTRACT

Body weight and nine morphostructural characters (withers height, rump height, heart girth, body length, head width, cannon circumference, shoulder width, rump width and rump length) of 83 White Fulani cows aged 1.5-2.4 years old were used to study the problem of multicollinearity instability in the estimation of body weight from morphological indices. Pairwise phenotypic correlations indicated a high and positive significant relationship between body weight and body dimensions (r = 0.61- 0.94; P<0.01). Among the linear type traits, the highest correlation was observed between withers height and rump height (r =0.98) while the lowest value was recorded for rump height and shoulder width (r =0.51). Severe collinearity problems were evident in 5 of the zoometrical variables as portrayed by variance inflation factors (VIFs) higher than 10.00 (VIF = 33.096, 31.421, 24.612, 22.726 and 13.327 for rump height, withers height, rump length, heart girth and body length respectively). Collinearity problems were further confirmed from the computations of the eigenvalues of the correlation matrix, condition indexes and variance proportions. Heart girth was retained among the collinear variables, and singly accounted for 87.9%, 92.3% and 94.1% of the variation in body weight in the subsequent stepwise regression, quadratic and cubic models, respectively.

Keywords: body weight/body measurement/multicollinearity/White Fulani cow/ stepwise regression

INTRODUCTION

In any regression analysis, the partial regression coefficients and partial sums of square for any independent variable are dependent on which other independent variables are in the model [10]. Inferences based on ordinary least squares regression can be influenced strongly by multicollinearity, and the fitted model, hence, may reflect unusual features because of the overall relationship among the variables [2]. Multicollinearity is defined as the existence of nearly linear dependency among columns of the design matrix X in linear prediction models. It induces numerical instability into the estimates and limits the size of the coefficient of determination. It also makes determining the contribution of each explanatory variable difficult because the effects of these variables are mixed. Regression coefficients may have the wrong sign (±) or an implausible magnitude. Accordingly, the partial regression coefficients are unstable and unreliable [8,13].

Cattle are the single most important livestock species in Nigeria in terms of animal protein, value and biomass [20]. The White Fulani are the most numerous of the Nigerian cattle breeds and have socio-economic importance and wide distribution in West African countries. According to RIM [14], the White Fulani represent about 37.1% of the national cattle population of Nigeria. Knowing the body weight of cattle is important for a number of reasons, related to selection, breeding, feeding and health care. However, this fundamental knowledge is often unavailable in the small scale farming sector, due to unavailability of scales. Although body measurements have been used in animals to predict body weight [9,17,21], information on the problem of multicollinearity among the predictors (body dimensions) is still scanty.

Therefore, the present investigation aimed at addressing the problem of multicollinearity in the prediction of body weight from morphometric traits of White Fulani cows.

MATERIALS AND METHODS

Experimental animals and their management

Data were obtained from eighty three White Fulani cows semi-intensively managed at the Livestock Complex, College of Agriculture, Lafia, Nasarawa State, North Central Nigeria. The animals were of two age categories: 1.5-1.9 years and 2.0-2.4 years, respectively. Age was determined from the available records on cows; and where information was missing its age was estimated using dentition. The Farm is located on Latitude 08^0 35'N and Longtitude 08^0 33'E respectively.

Parameters measured

Body weight (BW) and nine biometric traits were taken on each animal. The body parts measured were withers height (WH), rump height (RH), heart girth (HG), body length (BL), head width (HW), cannon circumference (CC), shoulder width (SW), rump width (RW) and rump length (RL).The anatomical points of reference have been previously described [23]. BW estimation was done using a scale. The height measurements were obtained using a graduated measuring stick. The length and circumference measurements were carried out using a measuring tape while the width measurement was done using a calibrated wooden calliper. All measurements were done by the same person in order to avoid between-individual variations.

Statistical analysis

Data (pooled for the two age categories) were analysed for preliminary descriptive statistics (Mean± SD, coefficient of variation, and minimum and maximum values). As a first indication of severity of collinearity, correlation coefficients among all the nine independent body measurements were estimated. Due to the inadequacy of correlation as a method of detecting collinearity, the method of variance inflation factor [15] was employed as follows:

$$VIF = \frac{1}{1 - R_i^2}$$

where,

R_i^2 = coefficient of determination.

Eigenvalues of the correlation matrix (X'X), condition indexes and variance proportions were also computed to confirm the existence or otherwise of collinearity following the procedures adopted by [10] and [13].

The following model described by [22] was employed to delete redundant variables arising from multicollinearity:

$$RV = | B_j | / б$$

where,

RV = redundant variable.

B_j = regression coefficient of X_j variable.

б = square root of residual mean square of the full regression model.

The full regression model (all the nine body measurements inclusive) was defined as:

$$Y = a + b_1X_1 + b_2X_2 + b_3X_3 + \text{------------------} + b_pX_p$$

where,

Y = dependent or endogenous variable (body weight)

a = intercept

b 's = regression coefficients

X's = independent or exogenous variables (WH, RH, BL, HG, HW, CC, SW, RW and RL)

The eventual regression models were fitted using stepwise multiple regression analysis. The quadratic and cubic effects of the predictors were also considered. Each model was assessed using R^2, adjusted R^2 and RMSE (Root mean squares error). SPSS [18] statistical package was employed in the analysis.

RESULTS AND DISCUSSION

Morphostructural traits

Mean (±SD) and coefficient of variation of each BW and biometric measurement of White Fulani cows are presented in Table 1. BW (kg), WH, RH, HG, BL, HW, CC, SW, RW, and RL (cm) averaged 116.01, 83.72, 89.9, 92.25, 110.93, 12.71, 12.05, 18.97, 20.65 and 27.68 respectively. The high variability in BW, HG, SW and RW could be attributed to certain environmental influence such as temperature and nutrition on these parameters.

Pairwise correlations

Bivariate correlations among BW and body dimensions of White Fulani cows are shown in Table 2. BW was positively and highly associated with morphostructural traits (r =0.61-0.94; P<0.05). Among the linear type traits, the highest correlation was observed between WH and RH (r =0.98) while the lowest estimate (r =0.51) was recorded for RH and SW. Similar findings have been reported [10, 12, 24]. The strong relationship existing between BW and body measurements suggests that either or combination of these morphological traits could be used to estimate live weight in cattle fairly well in the situation where weighbridges or scales are not available.

Detecting multicollinearity

The variance inflation factors (VIFs) and tolerance (T) values for the relationships among body dimensions of White Fulani cows are presented in Table 3. A bivariate correlation matrix of explanatory variables might not be sufficient to identify collinearity problems because near linear dependencies may exist among more complex combinations of regressors [4]. According to Weisberg

Table 1: Descriptive statistics of body weight (kg) and body dimensions (cm) of White Fulani cows

Trait	Mean (±SD)	CV	Minimum value	Maximum value
BW	116.01±31.18	26.88	75.00	240.00
WH	83.72±14.57	17.40	60.90	110.00
RH	89.90±16.23	18.05	64.40	115.00
HG	92.25 ±21.32	23.11	55.00	146.00
BL	110.93 ±28.27	25.48	71.00	178.20
HW	12.71 ±1.81	14.24	8.00	17.90
CC	12.05 ± 2.11	17.51	7.00	17.50
SW	18.97± 4.51	23.77	11.70	35.00
RW	20.65 ± 5.01	24.26	11.40	34.30
RL	27.68 ± 5.38	19.44	18.00	38.00

Table 2: Phenotypic correlations of body weight and biometric traits of White Fulani cows*

Trait	BW	WH	RH	HG	BL	HW	CC	SW	RW	RL
BW	-	0.77	0.76	0.94	0.89	0.61	0.73	0.79	0.82	0.87
WH	-	-	0.98	0.86	0.81	0.68	0.72	0.55	0.71	0.90
RH	-	-	-	0.86	0.81	0.63	0.74	0.51	0.68	0.90
HG	-	-	-	-	0.94	0.65	0.81	0.77	0.84	0.96
BL	-	-	-	-	-	0.64	0.84	0.72	0.76	0.94
HW	-	-	-	-	-	-	0.69	0.63	0.62	0.72
CC	-	-	-	-	-	-	-	0.64	0.70	0.84
SW	-	-	-	-	-	-	-	-	0.89	0.69
RW	-	-	-	-	-	-	-	-	-	0.78

*Significant at P<0.01 for all correlations

Table 3: Parameter estimates and variance inflation factors (VIF) of body measurements for estimating body weight in White Fulani cattle

Trait	Estimate	S.E.	Significance	R^2	VIF	Remarks
Intercept	0.360	9.462	0.970	-	-	-
WH	0.331	0.443	0.457	0.96	24.612	Collinearity
RH	-0.403	0.408	0.326	0.97	33.096	Collinearity
HG	1.390	0.257	0.001	0.96	22.726	Collinearity
BL	0.209	0.149	0.164	0.92	13.327	Collinearity
HW	0.122	1.147	0.916	0.69	3.244	Non-collinearity
CC	-1.621	1.162	0.167	0.78	4.544	Non-collinearity
SW	0.915	0.692	0.190	0.86	7.337	Non-collinearity
RW	-0.183	0.634	0.774	0.87	7.642	Non-collinearity
RL	-0.824	1.062	0.440	0.96	24.612	Collinearity

Table 4: Eigenvalues, condition indexes (CI) and variance proportions of body measurements for predicting body weight in White Fulani cows

Component	Eigen values	CI	C	WH	RH	HG	BL	HW	CC	SW	RW	RL
1	9.888	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
2	0.047	14.47	0.13	0.00	0.00	0.00	0.01	0.01	0.00	0.02	0.02	0.00
3	0.031	17.79	0.04	0.00	0.00	0.00	0.01	0.01	0.00	0.09	0.03	0.00
4	0.012	28.36	0.00	0.01	0.01	0.00	0.07	0.01	0.14	0.00	0.10	0.00
5	0.008	36.07	0.65	0.00	0.00	0.01	0.02	0.45	0.00	0.00	0.01	0.00
6	0.006	39.35	0.06	0.00	0.00	0.01	0.06	0.02	0.51	0.12	0.25	0.00
7	0.003	56.19	0.11	0.02	0.03	0.00	0.08	0.21	0.20	0.68	0.49	0.02
8	0.002	67.30	0.00	0.03	0.00	0.54	0.65	0.00	0.01	0.00	0.08	0.02
9	0.001	93.83	0.00	0.00	0.00	0.44	0.06	0.19	0.00	0.09	0.00	0.92
10	0.000	143.27	0.00	0.93	0.94	0.03	0.04	0.10	0.13	0.01	0.02	0.00

C: Constant

Table 5: Regression equations for the prediction of body weight from morphometric traits of White Fulani cows

Model		R^2	adj. R^2	RMSE
Stepwise				
BW=	$-10.456 + 1.371HG$	0.879	0.877	10.930
BW=	$-3.368 + 1.494HG - 1.533CC$	0.882	0.879	10.831
BW=	$-5.970 + 1.484HG - 1.708CC + 0.449HW$	0.883	0.878	10.883
BW=	$-4.567 + 1.379HG + 0.605RW - 1.589CC$	0.885	0.881	10.771
BW=	$-15.042 + 1.185HG + 1.144SW$	0.890	0.887	10.476
BW=	$-7.854 + 1.355HG - 0.479RW - 1.593CC + 1.475SW$	0.895	0.889	10.368
BW=	$-7.658 + 1.312HG - 1.615CC + 1.164SW$	0.894	0.890	10.347
Quadratic				
BW=	$111.734 - 1.303HG + 0.014HG^2$	0.923	0.921	8.770
Cubic				
BW=	$-142.156 + 6.974HG - 0.073HG^2 + 0.000HG^3$	0.941	0.938	7.745

[22], collinearity leads to large variances for estimated coefficients between variables. This informs the use of the VIF, which represents the increase in variance due to high correlation between the predictors. The severity of multicollinearity, however should not be quantified solely by the magnitude of pairwise correlations because the interrelation among three or more variables might result in a high degree of multicollinearity, even when pairwise correlations are low [16]. In the present study, the VIFs gave the first indication of the existence of severe collinearity in RH, WH, RL, HG and BL (VIF equals 33.096, 31.421, 24.612, 22.726, and 13.327, respectively). According to Gill [7], no absolute standard exists for judging the magnitude of the VIF. However, a crude rule of thumb is to be suspicious of collinearity if VIF is greater than 10.00.This is consistent with the report of Rook et al. [15].

The eigenvalues of the correlation matrix, condition indexes and variance proportions of the estimates further confirmed the problem of multicollinearity (Table 4). A close examination of this table revealed that there were three relatively small eigenvalues of 0.002, 0.001, and 0.000 for components 8, 9, and 10 respectively, showing how much the correlation matrix approached singularity. These components with small eigenvalues had large variance proportions of 0.50, 0.65, 0.92, 0.93 and 0.94 for HG, BL, RL, WH, and RH, respectively. The corresponding condition indexes were 67.300 (HG and BL), 93.833 (RL) and 143.269 (WH and RH). According to Malau-Aduli et al. [10], when trying to diagnose the reason for collinearity, the focus is on the principal components with very small eigenvalues because variables in multicollinearity are identifiable by their relatively large variance proportions with small eigenvalues. The variance proportions indicate the relative contribution from each principal component to the variance of each regression coefficient. The larger the condition index, the more the tendency towards collinearity. Belsley [4] suggested that moderate to strong relations are associated with condition numbers of 30 to 100.

Deletion of redundant variables

Collinearity implies that the effect of one predictor cannot be uniquely identified (i.e., is nearly confounded with the effect of another predictor). In such instance, the statistical model can include only one of the two predictors [19]. The deletion of one or more collinear variables improves the accuracy and robustness of the prediction models. According to Weisberg [22], the deletion of variables with small $| B_j |/ б$ would be desirable. The values obtained in the present study for WH, RH, HG, BL, and RL were 0.032, 0.039, 0.133, 0.020, and 0.079, respectively. Thus, among the collinear variables, HG was retained for the subsequent regression analysis.

Regression models for the prediction of body weight

The regression models for estimating BW from body measurements of White Fulani cows are presented in Table 5. The stepwise regression models revealed that HG singly accounted for 87.9% of the variation in BW. The RMSE in this case was 10.930. The model involving HG and SW improved the efficiency of the prediction equations (R^2, adjusted R^2, and RMSE were 0.890, 0.887, and 10.476, respectively). A slight improvement was obtained from the model involving the combination of HG, CC and SW (R^2, adjusted R^2, and RMSE equals 0.894, 0.890, and 10.347, respectively). The prediction model was greatly improved when the quadratic effect was tested (R^2, adjusted R^2, and RMSE were 0.923, 0.921, and 8.770, respectively). However, the best model for estimating BW from HG was obtained using the cubic model. This was because both the R^2 (0.941) and adjusted R^2 (0.938) of this model were highest, while the RMSE (7.745) was lowest. BW or size in general has long been considered as a paradigm for quantitative inheritance. Body measurements can be used to accurately predict BW [26]. The present observation is consistent with the report of [6], where the prediction of BW of cattle from HG gave R^2 value of 0.97. Similarly, Nagy [11] reported that the model including HG and cannon girth gave a good estimate of the BW of cattle, while Bagui and Valdez [3] used the formulas based on HG (R^2 = 0.943) and combination of HG and SW (R^2 = 0.953) to predict BW of Brahman cattle. In a similar study on Azawak Zebu in Niger, Dodo et al. [5] accentuated the significance of HG as a predictor of BW. A high genetic relationship between BW and HG had also been reported by Afolayan [1], thereby justifying its use for selection purposes and weight estimation. The importance of HG in weight estimation could be as a result of the fact that the muscle and a little of fat along with bone structure contribute to its formation [25].

CONCLUSIONS

Bivariate correlations between BW and body dimensions of White Fulani cows were positive and highly significant. The problems of multicollinearity were evident in RH, WH, RL, HG, and BL as revealed by variance inflation factors, eigenvalues of the correlation matrix, condition indexes and variance proportions. Among the collinear variables, HG was retained and singly explained 87.9% variation in the BW of cows in the stepwise regression analysis. When HG was fitted in the quadratic model, the prediction equation greatly improved. However, BW was best predicted from HG when the cubic effect was tested. The practical implications of this study are that BW of cows can be fairly estimated in the field using biometric traits for selection purposes, feeding, health and as a way of estimating market values in terms of cost of animals.

REFERENCES

[1] Afolayan R.A., Genetics of growth and development in cattle. Ph.D Thesis, The University of Adelaide, SA, Australia, 2003.

[2] Aziz M.A., Sharaby, M.A., Collinearity as a problem in predicting body weight from body dimensions of Najdi sheep in Saudi Arabia. Small Rumin. Res. (1993) 12, 117-124.

[3] Bagui N.J.G., Valdez, C.A., Live weight estimation of locally raised adult purebred Brahman cattle using

external body measurements. Phillipine Journal of Veterinary Medicine (2007) 44, 36-42.

[4] Belsley D.A., Kuh, E. and Welsch, R.E., Regression diagnostics: Identifying influential data sources of collinearity. 1st Edition, John Wiley and Sons, New York, 1980, pp. 292.

[5] Dodo K., Pandey, V.S., Illiassou, M.S., Weight estimation from body measurements of the Azawak Zebu in Niger. Revue Elev. Med. Vet. Pays trop. (2001) 54, 63-68.

[6] Francis J., Sibanda, S., Hermansen, J.E., Kristensen, T., Estimating body weight of cattle using linear body measurements. Zimbabwe Veterinary Journal (2002) 33, 15-21.

[7] Gill J.L., Outliers and influence in multiple regression. J. Anim. Breed. Genet. (1986) 103, 161-175.

[8] Hair Jr, J.E., Anderson, R.E., Tatham, R. L., Black, W.C., Multivariate data analysis. 3rd Edition, Macmillan Publishing Company, New York, 1992, pp. 544.

[9] Heinrichs A.J., Rogers, G.W., Cooper, J.B., Predicting body weight and wither height in Holstein Heifers using body measurements. J. Dairy Sci. (1992) 75, 3576-3581.

[10] Malau-Aduli A.E.O., Aziz, M.A., Kojina, T., Niibayashi, T., Oshima, K. Komatsu, M., Fixing collinearity instability using principal component and ridge regression analyses in the relationship between body measurements and body weight in Japanese Black cattle. J. Anim. Vet. Adv. (2004) 3, 856-863.

[11] Nagy B., Some production traits of Hungarian Grey cattle. Ph.D Thesis (electronic version), Georgikon Faculty of Agriculture, University of Pannonnia, Keszthely, Hungary, 2007, pp.18.

[12] Nwacharo J.M., Okeyo, A.M., Kamande, G.K., Rege, J.E.O., The small East African shorthorn zebu cows in Kenya. 1: Linear body measurements. Trop. Anim. Health Prod. (2006) 38, 65-76.

[13] Pimentel E.C.G., Queiroz, S.A., Carvalheiro, R., Fries, L.A, Use of ridge regression for prediction of early growth performance in crossbred calves. Genet. Mol. Biol. (2007) 30, 536-544.

[14] RIM, Nigerian Livestock Resources 2: National Synthesis. Resource Inventory and Management Ltd., Jersey, UK, 1992.

[15] Rook A.J., Dhanoa, M.S., Gill, M., Prediction of the voluntary intake of grass silages by beef cattle. 2. Principal component and ridge regression analyses. Anim. Prod. (1990) 50, 439-454.

[16] Roso V.M., Schenkel, F.S., Miller, S.P., Schaeffer, L.R., Estimation of genetic effects in the presence of multicollinearity in multibreed beef cattle evaluation[1]. J. Anim. Sci. (2005) 83, 1788-1800.

[17] Sowande O.S., Sobola, O.S., Body measurements of West African Dwarf sheep as parameters for estimation of live weight. Trop. Anim.Health Prod. (2007) 40, 433-439.

[18] SPSS, Statistical Package for the Social Sciences. SPSS Inc., 444 Michigan Avenue, Chicago, IL60611, 2001.

[19] St-Pierre N.R., Meta-analyses of experimental data in the animal sciences. R. Bras. Zootec. (2007) 36, 343-358.

[20] Tewe O.O., Sustainability and Development: Paradigms from Nigeria Livestock Industry. University Inaugural Lecture delivered on October 9, 1997 on behalf of the Faculty of Agriculture and Forestry, University of Ibadan, Ibadan, 1998.

[21] Vargas C.A., Elzo, M.A., Chase, C.C., Olson, T.A., Genetic parameters and relationships between hip height and weight in Brahman cattle. J. Anim. Sci. (2000) 78: 3045-3052.

[22] Weisberg S., Applied linear regression. 2nd Edition, John Wiley and Sons Publishers, New York, USA, pp. 196, 1985.

[23] Yakubu A., Ogah D.M., Idahor, K.O., Principal component analysis of the morphostructural indices of White Fulani cattle. Trakia Journal of Sciences (2009) 7: 67-73.

[24] Yakubu A., Ogah, D.M., Ladokun, A.O., Adua, M.M., Comparison of linear and non-linear models in the estimation of body weight from body measurements in West African Dwarf goats in north central Nigeria. Proceedings 32nd Annual Conference of the Nigerian Society for Animal Production, Calabar, March 18-21,2007, pp. 130-133.

[25] Yakubu A., Path coefficient and path analysis of body weight and biometric traits in Yankasa lambs. Slovak J. Anim. Sci. (2010) 43, 17-25. [26] Yan T., Mayne C.S., Patterson, D.C., Agnew, R.E., Prediction of body weight and empty body composition using body size measurements in lactating dairy cows. Livestock Science (2009) 124, 233-241