



Multicollinearity among morphometric traits in Madgyal sheep

DINESH KUMAR YADAV¹, REENA ARORA² and ANAND JAIN³

ICAR-National Bureau of Animal Genetic Resources, Karnal, Haryana 132 001 India

Received: 13 September 2017; Accepted: 14 December 2017

Key words: Madgyal sheep, Morphometric traits, Multicollinearity, Variance inflation factor

Morphological characters provide useful information to detect genetic structure and individual breed's potentiality due to intrinsic relationship among all biological characters (Zaiton *et al.* 2005). Body measurements have been used in sheep to contrast variation in shape and size (Salako 2006, Yadav *et al.* 2016) and to estimate body weight (Topal and Macit 2004, Yadav and Bhar 2008). Phenotypic characterization of sheep focuses primarily on describing the linear type traits, which are generally used to predict the body weight using multiple regression models. In some cases, multiple regression results may seem paradoxical. The model may fit the data (high F-test), even though none of the independent variables (X) has a statistically significant impact of explaining the dependent variable (Y). When this happens, the X variables are collinear and the results show multicollinearity, an undesirable situation where the correlations among the independent variables are strong. Multicollinearity misleadingly inflates the standard errors and makes some variables statistically insignificant while they should be otherwise significant. It is commonly observed in economic variables. This paper investigates multicollinearity among morphometric traits of Madgyal sheep of Maharashtra, India.

Data on morphometric traits of 200 Madgyal sheep were recorded from Sangli district of Maharashtra, the habitat of Madgyal sheep (Yadav *et al.* 2015, 2016). The traits measured are given in Table 1. Multicollinearity was detected by examining the pair-wise correlation matrix of the predictors, variance inflation factor ($VIF=1/1-R^2$, where R^2 is the coefficient of determination), tolerance values ($T=1-R^2$) and eigenvalues of the correlation matrix, condition indices and variance proportions (Malau-Aduli *et al.* 2004, Pimentel *et al.* 2007). Statistical analysis was carried out using SAS (2012).

The simplest and most obvious means of identifying multicollinearity is the presence of high correlation (≥ 0.90) among the predictor variables. Phenotypic correlations between the morphometric traits of Madgyal sheep are given in Table 1. Body weight was best associated with chest girth followed by thorax depth and body length. High correlation

coefficient was observed between withers height and rump height (0.97) indicating presence of collinearity.

The use of several variables as predictors makes the assessment of multiple correlation between the independent variables necessary to identify multicollinearity but misses more subtle forms of multicollinearity. Correlation matrix shows only simple correlations between two variables but near linear dependencies may exist among more complex combinations of regressors (Pimentel *et al.* 2007). Variance inflation factor, another collinearity diagnostics, indicates whether a predictor has a strong linear relationship with the other predictor(s). Related to the VIF is the tolerance statistic, which is its reciprocal. Table 2 gives the variance inflation factors (VIFs) and tolerance (T) values among the relationships of body measurements of Madgyal sheep. The results indicated presence of severe collinearity in withers height and rump height. Although there are no hard and fast rules about what value of the VIF should cause concern, there are some general guidelines. If VIF is equal to 1, there is no multicollinearity among the predictors but if VIF is greater than 1, the predictors may be moderately correlated. Myers (1990) suggested that a value of 10 is good value at which to worry. Bowerman and O'Connell (1990) suggested that if the average VIF is greater than 1, then multicollinearity may be biasing the regression model. Tolerance below 0.1 indicates the serious problem. Tolerance below 0.2 indicates a potential problem (Menard 1995). A VIF between 5 and 10 indicates high correlation that may be problematic. And if VIF is greater than 10, it can be assumed that the regression coefficients are poorly estimated due to multicollinearity. Small tolerance values also confirmed collinearity problem (Table 2).

Other measures that are useful for assessing multicollinearity are eigenvalues of the correlation matrix, condition index and variance proportion of the estimates (Table 3). The condition index represents the collinearity of combinations of variables in the data set (actually the relative size of the eigenvalues of the matrix). Identify all condition indices above a threshold value of 30, the most commonly used value. For all condition indices exceeding the threshold value, identify variables with variance proportions above 90%. A collinearity problem is indicated for two or more coefficients satisfying these conditions.

Present address: ¹Principal Scientist (dkyadav66@gmail.com), ²Principal Scientist (rejagati@yahoo.co.in), ³Principal Scientist (anand_nbagr@rediffmail.com).

Table 1. Correlation coefficients among morphometric traits of Madgyal sheep

Trait	BW	BL	WH	RH	CG	EL	TL	TD	RW	RL	ShC	TC	HL	HW
BW	1.00	0.80	0.76	0.77	0.90	0.27	0.39	0.86	0.67	0.78	0.62	0.40	0.70	0.68
BL		1.00	0.88	0.87	0.71	0.37	0.52	0.75	0.60	0.78	0.65	0.38	0.75	0.59
WH			1.00	0.97	0.70	0.39	0.47	0.76	0.57	0.78	0.68	0.35	0.75	0.59
RH				1.00	0.69	0.37	0.47	0.77	0.57	0.79	0.71	0.38	0.75	0.62
CG					1.00	0.29	0.28	0.85	0.65	0.71	0.49	0.37	0.64	0.59
EL						1.00	0.33	0.29	0.30	0.29	0.21	0.07	0.39	0.20
TL							1.00	0.35	0.24	0.38	0.47	0.24	0.40	0.40
TD								1.00	0.68	0.75	0.58	0.38	0.63	0.65
RW									1.00	0.66	0.43	0.32	0.52	0.45
RL										1.00	0.67	0.36	0.67	0.64
ShC											1.00	0.41	0.59	0.62
TC												1.00	0.34	0.34
HL													1.00	0.62
HW														1.00

BW, body weight; BL, body length; WH, withers height; RH, rump height; CG, chest girth; EL, ear length; TL, tail length; TD, thorax depth; RW, rump width; RL, rump length; ShC, shin circumference; TC, tail circumference; HL, head length; HD, head width.

Table 2. Variance inflation factor (VIF) and tolerance (T) values for the morphometric traits (N=200)

Morphometric trait	VIF	T
BL	5.982	.167
WH	23.234	.043
RH	23.018	.043
CG	4.224	.237
EL	1.313	.761
TL	1.581	.632
TD	5.291	.189
RW	2.205	.454
RL	3.970	.252
ShC	2.560	.391
TC	1.314	.761
HL	2.936	.341
HW	2.307	.433

Our results (Table 3) indicated that coefficients of withers height and rump height were having condition index value (496) and variance proportion values (0.93 each) above the threshold values. While diagnosing the reason for collinearity, one target the principal components with very small eigen values because collinear variables are identified by their relatively large variance proportions with small eigenvalues (Malau-Aduli *et al.* 2004). The variance proportions indicates relative contribution from each principal component to the variance of each regression coefficient. The larger the condition index the more the tendency towards collinearity. Moderate to strong associations are indicated with condition index of 30 to 100 (Belsely *et al.* 1980). Similar results were shown by Yakubu (2009) in West African dwarf goats. If there are one or more predictors with a high VIF, one of the predictors should be removed from the regression model as they provide redundant information. Removing one of the collinear

Table 3. Eigenvalues, condition indices, and variance proportions of morphometric traits for prediction of body weights of Madgyal sheep

Component	Eigenvalue	Condition index	Variance proportion													
			(Constant)	BL	WH	RH	CG	EL	TL	TD	RW	RL	ShC	TC	HL	HW
1	13.964	1.000	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	
2	.012	34.181	.00	.00	.00	.00	.00	.01	.00	.00	.00	.00	.00	.84	.00	
3	.007	45.100	.02	.00	.00	.00	.00	.05	.00	.00	.00	.00	.06	.08	.34	
4	.005	52.668	.00	.00	.00	.00	.01	.01	.55	.01	.03	.00	.03	.00	.01	
5	.003	66.536	.00	.00	.00	.00	.00	.06	.04	.00	.00	.01	.43	.03	.41	
6	.002	78.380	.02	.00	.00	.00	.01	.42	.31	.01	.11	.00	.12	.01	.04	
7	.002	91.098	.04	.01	.01	.00	.00	.19	.01	.00	.44	.00	.14	.00	.00	
8	.001	101.538	.85	.01	.00	.00	.00	.15	.00	.00	.00	.00	.06	.00	.04	
9	.001	110.874	.01	.00	.00	.00	.14	.08	.00	.14	.23	.01	.03	.00	.25	
10	.001	124.692	.00	.01	.01	.01	.13	.01	.01	.01	.00	.23	.10	.01	.43	
11	.001	141.376	.00	.01	.02	.01	.07	.00	.02	.02	.14	.69	.00	.00	.04	
12	.000	184.901	.00	.20	.00	.00	.48	.00	.01	.64	.01	.05	.02	.00	.16	
13	.000	205.390	.01	.74	.04	.04	.15	.01	.05	.16	.03	.00	.01	.00	.00	
14	.000	496.040	.04	.01	.93	.93	.01	.00	.00	.00	.00	.01	.00	.02	.01	

predictors usually does not drastically reduce the R^2 value. The complete elimination of multicollinearity is not possible but the degree of multicollinearity can be reduced by adopting ridge regression and principal components regression (Malau-Aduli *et al.* 2004, Pimentel *et al.* 2007). The use of PCA factor scores which are orthogonal and not correlated eliminate the problem of collinearity. Yadav *et al.* (2016) demonstrated that the use of principal components was more appropriate than the use of original correlated variables in predicting body weight in sheep.

SUMMARY

Multicollinearity in regression is viewed as more of disadvantage, as it practically inflates unnecessarily the standard errors of the coefficients. Presence of multicollinearity among 13 morphometric traits of Madgyal sheep of Maharashtra was assessed. The multicollinearity diagnostics (pairwise correlation matrix, VIF, tolerance values, eigenvalues, condition indices and variance proportions) used in this study ascertained presence of collinearity between two morphometric traits (withers height and rump height) of Madgyal sheep. Besides revealing presence of collinearity among the morphometric traits, the study provides a good teaching example of collinear traits in domestic livestock.

REFERENCES

- Belsley D A, Kuh E and Welsch R E. 1980. Regression diagnostics: Identifying influential data sources of collinearity. 1st Edition, John Wiley and Sons, New York, pp. 292.
- Bowerman B L and O'Connell R T. 1990. Linear statistical models: An applied approach. 2nd ed. Duxbury, Belmont, CA, Malau-Aduli A E O, Aziz M A, Kojina T, Niibayashi T, Oshima K and Komatsu M. 2004. Fixing collinearity instability using principal component and ridge regression analyses in the relationship between body measurements and body weight in Japanese Black cattle. *Journal of Animal and Veterinary Advances* **3**: 856–63.
- Menard S. 1995. Applied logistic regression analysis. Sage University Paper Series on Quantitative Application in the Social Sciences, 07–106. Sage, Thousand Oaks, CA.
- Myres R. 1990. Classical and modern regression with applications. 2nd ed. Duxbury, Boston, MA.
- Pimentel E C G, Queiroz S A, Carvalheiro R and Fries L A. 2007. Use of ridge regression for prediction of early growth performance in crossbred calves. *Genetics and Molecular Biology* **30**: 536–44.
- Salako A E. 2006. Application of morphological indices in the assessment of type and function in sheep. *International Journal of Morphology* **24**: 13–18.
- Topal M and Macit M. 2004. Prediction of body weight from body measurements in Morkaraman sheep. *Journal of Applied Animal Research* **25**: 97–100.
- Yadav D K and Bhar L M. 2008. Robust model fitting in Muzaffarnagri sheep growth data under field conditions. *Indian Journal of Animal Sciences* **78**(11): 1285–87.
- Yadav D K, Arora R and Jain A. 2015. Sheep Genetic Resources of India: Madgyal Sheep. Monograph 91, ICAR-National Bureau of Animal Genetic Resources, Karnal, Haryana.
- Yadav D K, Arora R and Jain A. 2016. Principal component analysis of body measurements based morphological structure of Madgyal sheep. *Indian Journal of Animal Sciences* **86**(5): 568–71.
- Yakubu A. 2009. Fixing collinearity instability in the estimation of body weight from morpho-biometrical traits of West African dwarf goats. *Trakia Journal of Sciences* **7**(2): 61–66.
- Zaitoun I S, Tabbaa M J and Bdourb S. 2005. Differentiation of native goat breeds of Jordan on the basis of morphostructural characteristics. *Small Ruminant Research* **56**: 173–82.