Proceedings of the Pakistan Academy of Sciences:
A. Physical and Computational Sciences: 57 (2): 75-80 (2020)

ISSN: 2518-4245 (print), 2518-4253 (online)

Pakistan Academy of Sciences

Research Article

Application of CHAID Algorithm for the Identification of Morphological Traits of Indigenous Sheep Body Weight

Muhammad Arsalan[1*], Muhammad Aman Ullah[1], and Abdul Waheed[2]

[1]Department of Statistics, Bahauddin Zakariya University, Multan, Pakistan
[2]Department of Livestock and Poultry Production, Bahauddin Zakariya University, Multan, Pakistan

Abstract: The objective of this research study was the identification of significant morphological traits to predict the live body weight of indigenous sheep of southern Punjab, Pakistan. Application of Chi-Square Automatic Interaction Detection (CHAID) was used to achieve the objective of the current study. Pearson correlation technique was used to see the relationship of morphological traits with live body weight. The data of 13 morphological traits of 291 indigenous sheep was used using a simple random sample technique. The morphological traits such as barrel depth, body length, ear length, ear width, head length, head width, heart girth, neck length, neck width, rump length, rump width, tail length, and withers height were used. The data consists of 130 rams and 161 eves. The dependent variable was the live body weight of indigenous sheep. The barrel depth, rump width, and heart girth were strongly correlated with live bodyweight having correlations 0.968, 0.936, and 0.925 respectively. The result of CHAID analysis showed that barrel depth and heart girth are significant predictors of live body weight (***p-value*<0.001**).

Keywords: Body Weight, CHAID, Morphological trait, Predictor, Indigenous Sheep, Southern Punjab, Pakistan.

1. INTRODUCTION

The agriculture sector is the keystone of the economy of Pakistan. It is the second-largest sector and contributing about 21% of the national GDP. Livestock is contributing to the share of 58.92 percent to overall agriculture GDP. Livestock is the fastest growing sub-sectors in developing countries in the last few decades and it is still increasing continuously [16-17]. In Pakistan, especially in rural areas of southern Punjab, sheep are one of the important sources of income in form of meat and leather. The quantity of both meet and leather is directly proportional to sheep's live body weight. The sheep's live body weight is predicted through its body dimensions known as morphological traits. Many studies had been conducted to predict the live body weight of various breeds of sheep. Khan et al in 2014 [1] had conducted the studied to predict body weight through morphological traits as all of them had declared that bodyweight is highly correlated with morphological traits. [2] had also used various decision trees including CHAID to predict the live weight of Harnai sheep and found good. [3] used regression tree CART and CHAID for the selection of morphological traits that affect the continuous dependent variable weaning weight of Karayaka sheep breeding [4] had used regression tree method to determine the significant factors that influence the birth weight of Karakas and Norduz lamb in Turkey. Birth type, sex, damage, and genotype were found more significant factors in determining birth weight. [5] conducted a study to predict the bodyweight of Indigenous sheep breed in Balochistan, Pakistan using the regression tree method. The results of the study showed that the significant factors in determining body weight were withers height, body length, chest girth, and breed. A similar study was conducted by [6] on cattle breeds in Nigeria to access factors that influenced the prevalence of abortion. Except for the application of regression trees for determining significant predictors in biological studies; in many other fields, regression trees especially the CHAID algorithm was used in determining the significant factors. [7] used the regression tree methods to create

Received: February 2020; Accepted: June 2020
*Corresponding Author: Muhammad Arsalan <marsalan_s@yahoo.com>

predictive models in the field of engineering. [8] used CHAID analysis to determine socioeconomic variables that explain a student's academic success. [9] used regression tree methods in the study of determining happiness factors. The prior studies are solely related to the application of regression tree methods in determining the significant predictor of a nominal, ordinal, or continuous dependent variable without considering the distributional assumptions of variables.

2. MATERIALS AND METHODS

The data of 291 indigenous sheep from birth to 2 years of age were collected from Livestock Experiment Station Rakh Ghulaman and Rakh Khairewala. The data consists of measurement about following morphological traits of indigenous sheep i.e. live body weight, heart girth, body length, withers height, head length, head width, ear length, ear width, neck length, neck width, tail length, rump length, rump width, and barrel depth. In data, there were 130 rams and 161 eves. The dependent variable was live body weight. To determine the significant predictors of the live body, weight the zero-order Pearson correlations and CHAID algorithm were used in our analysis. CHAID chooses the predictor that has the strongest relationship with the dependent variable and categories of each predictor are merged if they are not statistically different concerning the dependent variable. Mohammad et al. [5] used Pearson correlation coefficients to check the bivariate relationships between morphological traits. Milanović [10] had focused on the importance of regression tree methods for the detection of data structure, especially in the multivariate data set. Koc et al. [11] declared the CHAID as the most suitable algorithm for the analysis of Animal Science data in his study as their study was based on lambs. Song [12] discussed in his study that the regression tree is a commonly used methodology for developing a prediction model for the target variable for large and complicated data set. The author had also discussed the non-parametric characteristic of the regression tree methods.

CHAID was originally introduced by [13] as an exploratory technique for investigating large quantities of categorical data. According to [13] the CHAID is an off-shoot of the AID (Automatic Interaction Detection) concept given by Morgon and Conquest in 1963. The basic theme is given by Kass (1980) in his study was to identify those subsets of predictors that best describe the dependent variable. Ritschard [14] had conducted a study to explain the functioning of CHAID. His interest was mainly to detect and account for nonlinear effects on the response variable. Çamdeviren and co-authors [15] had explained that the CHAID algorithm can also be applied to a continuous dependent variable by transforming into a categorical variable using cut-off values. The morphological traits were divided into barrel depth, body length, ear length, ear width, head length, head width, heart girth, neck length, neck width, rump length, rump width, tail length, and withers height. The authors also had explained that the regression tree method is not affected by a high correlation among predictors. The SPSS (SPSS Inc., Chicago, Illinois, USA) version 26.0 software was used for analysis in the study. The statistical significance was set for all statistical tests at a p-value < 0.05 (two-sided).

3. RESULTS AND DISCUSSIONS

Our analysis had been divided into two sections i.e. correlation analysis and CHAID analysis. The objective of both sections was to identify significant predictors of live weight.

3.1 Pearson Correlations

In this section, we had found zero-order correlations between each morphological trait (predictor) and live body weight. The results of correlations along with the significance value are given in Table 1. It was observed that barrel depth, rump width, and heart girth are highly correlated with live bodyweight having correlation coefficients 0.968, 0.936, and 0.925 respectively. These high correlations suggested the significant predictors of live body weight as the theoretical concept of linear regression and correlation.

3.2 CHAID Analysis

In this section, the CHAID algorithm had been constructed using SPSS software. The CHAID algorithm had been given in Figure 1. In the root node, dependent variable descriptive statistics had been given. All sheep had been included in the root node. The mean and standard deviation of live body

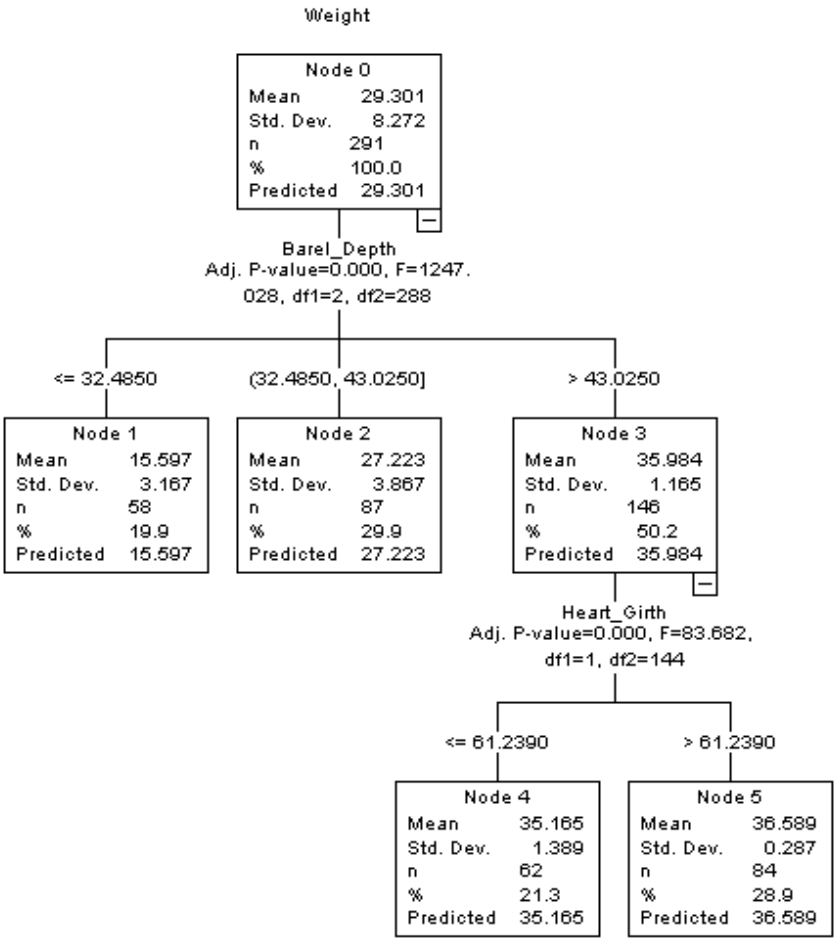

Fig.1. CHAID Algorithm

weight were 29.301and 8.272 respectively. The root node was divided into three child nodes 1, 2, and 3 by variable barrel depth with the first-degree effect on live body weight with significant p-value = 0.000.

The barrel depth is classified into three child nodes concerning its values. Child node 1 had been declared where barrel depth is less or equal to 32.4850. Child Node 2 had been declared in the range of barrel depth 32.4850 to 43.0250. Child Node 3 had been declared where barrel depth is greater than 43.0250. The number of sheep in child node 1, 2, and 3 were 58, 87, and 146 respectively. Mean live weight in child nodes 1, 2, and 3 were 15.597, 27.223, and 35.984 respectively while standard deviations were 3.167, 3.867, and 1.165 respectively. The child node 3 had more average weight and smaller standard deviation as compared to child node 1 and 2. Child nodes 1 and 2 were found to be terminal nodes while child node 3 was further divided into two-terminal nodes 4 and 5 by variable heart girth with the second-degree effect on live weight with significant p-value = 0.000. Terminal node 4 had been declared based on values of heart girth less than or equal to 61.2390 and terminal node 5 had the values of heart girth greater than 61.2390. The number of sheep in terminal node 4 and 5 were 62 and 84 respectively. Mean values of live weight in terminal nodes 4 and 5 were 35.165 and 36.589 respectively while standard deviations were 1.389 and 0.287 respectively. Node 0 is the root node, root 3 is the child node and root 1, 2, 4, and 5 are terminal nodes in our analysis. This algorithm showed the barrel depth and heart girth as the significant predictor of live weight.

Table 1 Pearson Correlation of traits with live weight

Sr. No	Morphological Traits	Pearson Correlation	
		Value (r)	Sig.
1	Barel Depth	0.968	0.000
2	Body Length	0.770	0.000
3	Ear Length	0.599	0.000
4	Ear Width	0.431	0.000
5	Head Length	0.924	0.000
6	Head Width	0.874	0.000
7	Heart Girth	0.925	0.000
8	Neck Length	0.337	0.000
9	Neck Width	0.577	0.000
10	Rump Length	0.909	0.000
11	Rump Width	0.936	0.000
12	Tail Length	0.570	0.000
13	Withers Height	0.796	0.000

4. CONCLUSIONS

The findings of the study by Pearson correlations and CHAID algorithm had shown that barely depth and heart girth are significant morphological traits, among all those included in the study, to predict the live body weight of indigenous sheep of southern Punjab. These two morphological traits had significant correlations with live weight and barrel depth had the first-degree effect on live weight while heart girth had a second degree effect on live weight. The major part of live weight depends upon the abdomen size of sheep while barrel depth and heart girth are the measurements related to abdomen size.

5. REFERENCES

1. M.A. Khan., M.M. Tariq, E. Eyduran, A. Tatliyer, M. Rafeeq, F. Abbas, N. Rashid, M.A. Awan, and K. Javed. Estimating body weight from several body measurements in Harnai sheep without a multicollinearity problem. *The Journal of Animal & Plant Sciences*, 24(1): 2014, Page: 120-126. ISSN: 1018-7081 (2014).
2. M. Ali., E. Eyduran, M.M. Tariq, C. Tirink, F. Abbas, M. A. Bajwa, M. H. Baloch, A. H. Nizamani, A. Waheed, M. A. Awan, S. H. Shah, Z. Ahmad, and S. Jan. Comparison of Artificial Neural Network and Decision Tree Algorithms used for Predicting Live Weight at Post Weaning Period from Some Biometrical Characteristics in Harnai Sheep. *Pakistan Journal of Zoology*. 47(6) 1579-1585 (2015).
3. M. Olfaz, C. Tirink, and H. Önder. Use of CART and CHAID Algorithms in Karayaka Sheep Breeding. *Kafkas Univ Vet Fak Derg* 25 (1): 105-110 (2019).
4. E. Eyduran., K. Karakus, S. Keskin, and F. Cengiz. Determination of Factors Influencing Birth Weight Using Regression Tree (RT) Method. *Journal of Applied Animal Research*, 34:2, 109-112 (2008).
5. M.T. Mohammad., M. Rafeeq, M. A. Bajwa, M. A. Awan, F. Abbas, A. Waheed, F. A. Bukhari, and P. Akhtar. Prediction of Body Weight from Body Measurements Using Regression Tree (RT) Method for Indigenous Sheep Breeds in Balochistan, Pakistan. *The Journal of Animal & Plant Sciences,* 22(1): 2012, Page: 20-24. ISSN: 1018 – 7081. (2012).
6. A. Yakubu., A. D. Awuje, and J. N. Omeje. Comparison of multivariate logistic regression and classification Tree to assess factors influencing the prevalence of abortion in Nigerian cattle breeds. *The Journal of Animal & Plant Sciences,* 25(6): 2015, Page: 1520-1526 (2015).
7. T. Fehér. Using Regression Trees in Predictive Modeling. *Production Systems and Information Engineering Volume* 4 (2006), pp. 115-124 (2006).
8. E. Önder, and Ş. Uyar. CHAID Analysis to Determine Socioeconomic Variables that Explain Students' Academic Success. *Universal Journal of Educational Research* 5(4): 608-619, (2017).
9. Y.B. Yücel. Determination of factors affecting happiness level by Classification tree technique. *European Journal of Business and Social Sciences*. 6(2) 54-62 (2017).
10. M. Milanović, and M. Stamenković. CHAID Decision Tree: Methodological Frame and Application. *Economic Themes* (2016) 54(4): 563-586 (2016).
11. Koc, Y., E. Eyduran, and O. Akbulut. Application of Regression Tree Method for Different Data from Animal Science. *Pakistan Journal of Zoology*. 49(2) 599-607 (2017).
12. Y.Y. Song, and Y. Lu. Decision tree methods: applications for classification and prediction. *Shanghai Archives of Psychiatry*. 27(2) 130-135 (2015).
13. G.V. Kass. An Exploratory Technique for Investigating Large Quantities of Categorical Data. *Applied Statistics,* 29(2) 119-127 (1980).
14. G. Ritschard. CHAID and Earlier Supervised Tree

Methods. In J.J. McArdle & G. Ritschard (eds), Contemporary Issues in Exploratory Data Mining in Behavioral Sciences, Routledge, New York, pages 48-74 (2013).

15. H. Camdeviren., M. Mendes, M.M. Ozkan, F. Toros, T. Sasmaz, and S. Oner. Determination of depression risk factors in children and adolescents by regression tree methodology. Acta Med. Okayama, 2005 Vol. 59, No. 1, pp. 19-26 (2005).
16. Pakistan Economic Survey (2018-19) http://www.finance.gov.pk/survey/chapters_19/2-Agriculture.pdf.
17. M. Qasim, M. Amin, M. N. Akram, T. Omer, and F. Hussain, Forecasting Buffalo Population of Pakistan using Autoregressive Integrated Moving Average (ARIMA) Time Series Models. *Proceedings of the Pakistan Academy of Sciences: A. Physical and Computational Sciences,* 56(3), 27-36 (2019).