Acta Agronomica Óváriensis

J., Tőzsér et al, /ActAgrOvar, Vol.65.1. (2024)

DETERMINING THE TYPE OF CHAROLAIS COWS BY CHAID ANALYSIS

JÁNOS TŐZSÉR[1], BENCE TARR[2]*, NATASA FAZEKAS[3], ZOLTÁN DOMOKOS[4]

[1]Albert Kázmér Faculty of Agricultural and Food Sciences of Széchenyi István University
Department of Animal Science, Mosonmagyaróvár, Hungary
ORCID: https://orcid.org/0000-0002-5632-1765
[2]Hungarian University of Agriculture and Life Sciences, Szent István Campus
Institute of Technical Sciences, Gödöllő, Hungary
email: tarr.bence.gyula@uni-mate.hu, ORCID: https://orcid.org/0009-0004-1790-9234
[3]Hungarian University of Agriculture and Life Sciences, Szent István Campus
Institute of Animal Husbandry, Gödöllő, Hungary
ORCID: https://orcid.org/0009-0002-4184-4616
[4]Beef cattle advisor/former managing director of Hungarian Charolais Breeders Association
Miskolc, Hungary
ORCID: https://orcid.org/0000-0001-5229-0212

*Levelező szerző/corresponding author

Érkezett/Received: 2024.01.16. Átdolgozva/Revised: 2024.02.19. Elfogadva/Accepted: 2024.03.05.

ABSTRACT

Factors influencing diagonal body length of Charolais cows (n=311) were analysed by the CHAID (Chi-squared Automatic Interaction Detector) method and are presented in this paper. CHAID algorithm groups the dependent variables (Y) so that the variance within groups is as low as possible and the variance between groups is as high as possible. The hierarchy of explanatory variables in this study was shown by the order of live weight, height at withers and age of the cows. The 1st, 2nd and 3rd nodus were formed by the live weight, according to the following categories: ≤600 kg, 600-645 kg, >645 kg. Diagonal body length increased parallel with the live weight categories. The 158 cows forming the first nodus were divided into two separate groups by the program, based on height at withers: 4th nodus (n=50, ≤ 128 cm); 5th nodus (n=108, >128 cm). Finally, the 5th nodus was separated – according to the age – into two groups in the last phase of the analysis: 6th nodus (n=53, ≤ 5.35 age, 172 cm); 7th nodus (n=55, > 5.35 age, 177 cm). Grouping individuals within the population would help to identify cow types, therefore CHAID analysis may be a possible useful method in beef cattle breeding.

Keywords: CHAID, Charolais, beef cattle, body measurements, body length

DOI: 10.17108/ActAgrOvar.2024.65.1.114

INTRODUCTION

A huge number of data are generated everywhere and new ways of understanding the data are developed for data analysts. One of the solutions for understanding the relationships of variables in a big dataset is the decision tree- also known as classification tree. This method helps in defining decision rules, segmenting and classifying data during solving a given professional task.

The main function of decision trees is to segment a complex decision problem into some smaller elements. The method itself is a decision-preparing tool, which describes the whole decision situation. The tree shows the consequences of choosing different solutions (Winston, 2003).

There are complex statistical algorithms behind the trees, one of them is CHAID (Chi-squared Automatic Interaction Detector). CHAID analysis is a multivariate recursive classification method designed by Kass (1980).

The process was originally designed for ordinal variables, but an advanced version of the algorithm is suitable for analysing continuous numeric dependent and independent variables (Hámori, 2001).

The main task of the algorithm is to group the dependent variables (Y) in such a way that the variance within groups is kept as low as possible, while the variance between groups is the highest. The hierarchy of explaining variables is determined by the analysis, based on their proportion to the variance of the dependent variable (Magidson, 1993).

CHAID is a popular segmenting method because: it is considered to be a better method than traditional cluster analysis, as the latter one can deal only with quantitative variables.

The reason this method is becoming quickly widespread and popular is, that it shows the relationship matrix of a given dependent variable and the possible explaining variable in a visual, easily acceptable decision/classification tree form.

It can be used for any ordinal or numeric data, without any restriction on its distribution, which is beneficial for users (data miners) (Jánosa, 2011)

There's still a limit in the method, as when integrating categories, it doesn't always reach the category structure when p_x is the smallest, so the distribution of the categories is optimal. As a consequence, the uniting algorithm stops and the remaining category pairs are considered statistically independent by the program.

A renewed version of the original method developed by Biggs, De Ville, & Suen (1991) – called Exhaustive CHAID – tried to solve that problem. However, in practice, CHAID and Exhaustive CHAID frequently result in the same tree structure for a given task.

CHAID has recently been applied in livestock breeding. The majority of such studies deal with dairy cattle: some papers reported the hierarchy of production traits and environmental factors affecting milk yield (Akilli, Atil, & Takma, 2022; Bakir, Keskin, & Mirtagloglu, 2009) birth weight (Yilmaz, Eyduran, & Kaygisiz, 2013), mastitis (Alaty, Aytelin, & Eyduran, 2022). and somatic cell count (Sahin, 2023). However, studies with this statistical method on beef cattle are limited, (Aksoy, Ertürk, Eyduran, & Tariq, 2018) and deal with the beef industry enterprise level, investigating the effect of socio-economic factors of the farmers on fattening live weight next to some traits of the animals (sex, starting weight, fattening period). Body measurements, exactly height at withers, sacrum

DOI: 10.17108/ActAgrOvar.2024.65.1.114

height, chest depth, width and circumference – and their effect on live weight – were analysed with CHAID method on sheep as well (Tirink, Piwczynski, Kolenda, & Önder, 2023).
Domokos & Tőzsér (2014) introduced the physiological reasons for the necessarily developed different types (breeder, butcher, intermediate and ranch) within the Charolais breed, and also the selection methods for calving ease, the change in viewpoint in French breeding organisation and evaluated the types in Hungary.
Based on literature data, Domokos (2011) highlighted, that the type differentiation of animals nowadays is also supported by linear conformation scores and digital image analysis methods in addition to body measurements. He emphasised the importance of body fat content (condition) as alterations in it led to important reproduction and production changes and at the same time, the cow's condition when giving birth determines the viability of the calf and the possibility of the next fertilisation.
Based on the Hungarian Charolais database, Domokos (2011) confirmed that there are strong positive correlations between thigh length, chest depth, pin width, muzzle width and cannon width. Based on those results he suggested decreasing the number of traits evaluated during type classification.
The aim of the present study was to reveal using CHAID analysis what other traits determine diagonal body length in a Charolais breeding station.

MATERIALS AND METHODS

The raw data of a previously published paper (Tőzsér, Domokos, Alföldi, Holló, & Rusznák, 2001). were used for the analysis. The measurements were taken on a Charolais farm in 2001, in the beginning of the summer when the animals were grazing. The cows (n= 311) had an average age of 6.8 years and an average weight of 600 kg. The traditional equipment was used to record body measurements (measuring stick and tape) and following Horn's (1976) suggestions the following measurements were recorded in addition to the weight on the scale: height at withers (132.2±3.93 cm), thigh width 2 (52.1,2±2.74 cm), diagonal body length (177,2±8,09 cm), chest circumference (194.5±8.50 cm).

The steps of the recursive algorithm forming the CHAID model are as follows:
- Defining the dependent variable.
- Uniting the statistically least correlating categories in case of every explaining variable (merging).
- Dividing the categories of the least independent explaining variable for the dependent variable (splitting).

The algorithm continues merging and splitting recursively until reaches a pre-defined end criterion (stopping).

DOI: 10.17108/ActAgrOvar.2024.65.1.114

Specifications of the analysis are shown in *Table 1*.

Table 1 Characteristics of the CHAID model

Model Summary		
Specifications	Growing Method	CHAID
	Dependent Variable	Diagonal body length, cm
	Independent Variables	Age Live weight, kg Height at withers, cm Chest circumference, cm Thigh width 2, cm
	Validation	Cross Validation
	Maximum Tree Depth	3
	Minimum Cases in Parent Node	100
	Minimum Cases in Child Node	50

Statistical analysis was performed using the SPSS 24 (SPSS Inc., Chicago, IL) statistical software.

RESULTS

After assessment only three parameters (live weight, height at withers, age) remained out of the original five independent variables when the number of nodes was 8, the number of terminal nodes was 5 and the depth of the tree was 3.

The diagonal body length is one of the important body measurements because more Charolais breeders try to improve this conformation trait mainly in the breed type. According to practical observations, increasing diagonal body length has an effect on meat production and ease of calving. Diagonal body length showed a moderate correlation with live weight ($r=0.63$, $P<0.05$), while its correlation with age was loose ($r=0.18$, $P<0.05$). Live weight is the most determinant parameter as shown in *Figure 1*, when 1st, 2nd and 3rd nodus were formed by live weight, according to the following categories: ≤600 kg, 600-645 kg, >645 kg. Diagonal body length increased parallel with the live weight categories: 173,2 cm, 179,5 cm, 184 cm.

The 158 cows making up the first nodus were divided into two separate groups by the program, based on height at withers: 4th nodus (n=50, ≤ 128 cm); 5th nodus (n=108, >128 cm). The correlation between height at withers and diagonal body length was $r= 0.54$ ($P<0.05$). Diagonal body length was 169,6 cm and 174,9 cm, respectively, in the two nodes.

DOI: 10.17108/ActAgrOvar.2024.65.1.114

The 5th nodus was separated – according to age – in two groups in the last phase of the analysis: 6th nodus (n=53, ≤ 5.35 age, 172,6 cm); 7th nodus (n=55, > 5.35 age, 177 cm).
The hierarchy of explaining variables was live weight, height at withers and age in this analysis. It's an interesting result that the group of cows with live weight ≤600 kg was further divided by height at withers. In the case of the cow group with height at withers over 128 cm (n=108) (5th nodus) age also had an influence on diagonal body length (*Figure 1*).

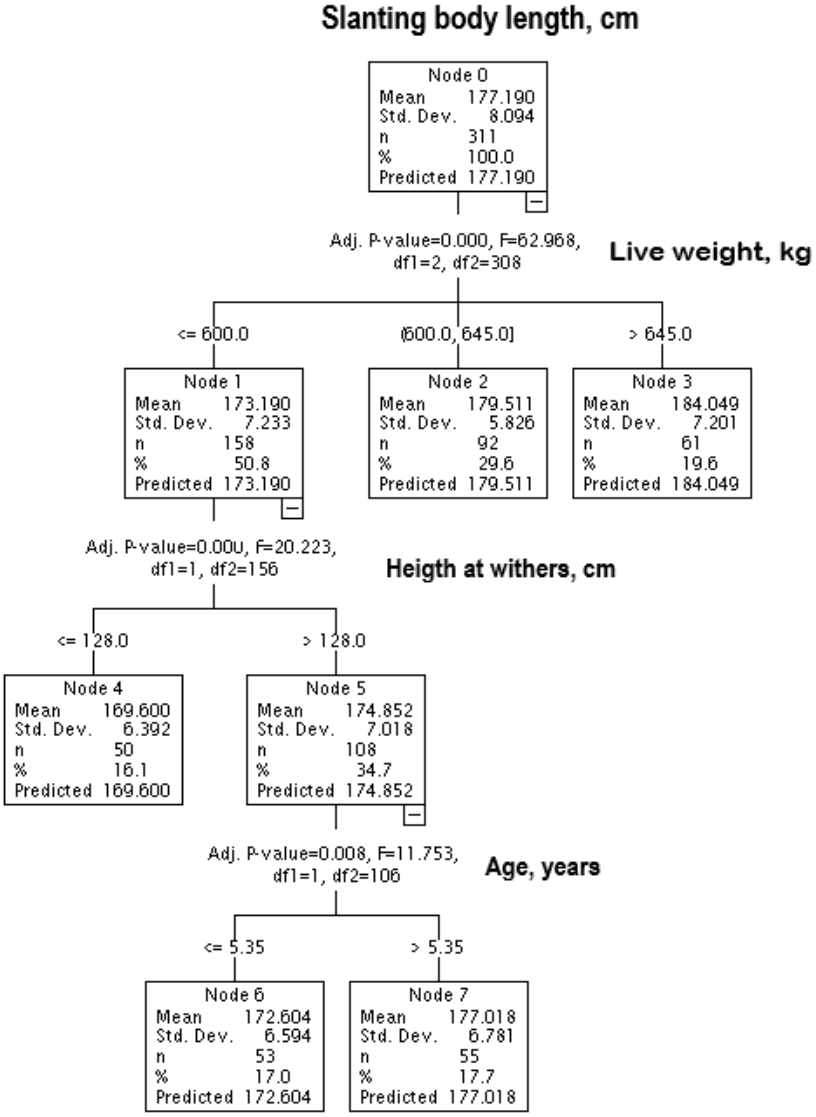

Figure 1 Decision tree on diagonal body length for Charolais cows

DOI: 10.17108/ActAgrOvar.2024.65.1.114

DISCUSSION AND CONCLUSION

Relationships and mutual connections of important traits are relevant information for successful selection in beef cattle breeding when deciding about selection criteria and defining mating plan.
At this moment the beef breeders use Excel software or special herd software to classify/group their herd. It is known that the CHAID can be used for analysing any ordinal or numerical data, so the use of this method can be useful for beef breeders. Hierarchical representation of independent variables enables the user to quickly interpret the results. Traditional procedures cannot perform hierarchical grouping.
We suppose that if the statistical program (such as SPSS) is available on a server, then the analysis can be done on the spot even with a tablet. Considering the rapid development of information technology, we expect that CHAID program will also be available with a smartphone shortly.
In our study, the CHAID procedure for diagonal body length showed how certain parameters (live weight, height at withers and age) were related to each other, and classified the cows into homogenous groups. Grouping of individuals helps the breeder to identify the different breed types. The application of CHAID analysis seems to be a good solution for beef cattle breeding, as it gives a more differentiated picture of the connection between economically important traits than other statistical methods.

A CHAROLAIS TEHENEK TÍPUSÁNAK MEGHATÁROZÁSA CHAID-ELEMZÉSSEL

ÖSSZEFOGLALÁS

A Charolais tehenek (n=311) ferde törzshosszúságát befolyásoló tényezőket CHAID (Chi-squared Automatic Interaction Detector) módszerrel elemezték. A CHAID egy olyan többváltozós rekurzív osztályozási eljárás, amelyben az algoritmus csoportosítja a függő változókat (Y), azért, hogy a csoporton belüli variancia a lehető legalacsonyabb legyen, ugyanakkor a csoportok közötti variancia pedig a legmagasabb. A magyarázó változók hierarchiája az élősúly, a marmagasság és a tehenek életkora szerint alakult. Ebben a vizsgálatban az 1., 2. és 3. csomópontok az élősúly alapján számítódtak, a következő kategóriák szerint: ≤600 kg, 600-645 kg, >645 kg. A ferde törzshosszúság az élősúly-kategóriákkal párhuzamosan nőtt. Az első csomópontot alkotó 158 tehenet a program két külön csoportra osztotta marmagasság alapján: 4. csomópont (n=50, ≤ 128 cm); 5. csomópont (n=108, >128 cm). Az 5. csomópont – életkor szerint – két csoport szerint különült el az elemzés utolsó fázisában: 6. csomópont (n=53, ≤ 5,35 életkor, 172 cm); 7. csomópont (n=55, > 5,35 életkor, 177 cm). Az egyedek populáción belüli csoportosítása segítene a tehéntípusok felismerésében, így a CHAID analízis egy lehetséges módszernek bizonyult a húsmarha-tenyésztésben.
Kulcsszavak: CHAID, charolais fajta, húsmarha, testméretek, törzshossz

DOI: 10.17108/ActAgrOvar.2024.65.1.114

REFERENCES

Akilli, A., Atil, H., & Takma, C. (2022). A decision tree model to determine some environmental factor affecting 305-day milk in Simmental cows. *Journal of Agricultural Faculty of Gaziosmanpasa University (JAFAG), 39*(3), 191-196. http://doi.org/10.55507/gopzfd.1175502

Aksoy, A., Ertürk, E.Y, Eyduran, E. & Tariq, M.M. (2018). Comparing predictive performances of MARS and CHAID algorithms for defining factors affecting final fattening live weight in Cultural Beef Cattle Enterprises. *Pakistan J. Zool., 50*(6), 2279-2286. http://doi.org/10.17582/journal.pjz/2018.50.6.2279.2286

Alaty, Y., Aytelin, I., & Eyduran, E. (2022). Use of multivariate adaptive regression splines, classification tree and roc curve in diagnosis of subclinical mastitis in dairy cattle. *Journal of the Hellenic Veterinary Medical Society*, *73*(1), 3817-3826. https://doi.org/10.12681/jhvms.25864

Bakir, G., Keskin, S., & Mirtagloglu, H. (2009). Evaluating the relationship between mature age milk yield and several traits using CHAID analysis in Brown Swiss cows. *Journal of Animal and Veterinary Advances***,** *8*(3), 587-589. Retrieved from https://docsdrive.com/?pdf=medwelljournals/javaa/2009/587-589.pdf

Biggs, D., De Ville, B., & Suen, E. (1991). A method of choosing multiway partitions for classification and decision trees. *Journal of Applied Statistics, 18*(1), 49-62. http://doi.org/10.1080/02664769100000005

Domokos, Z. (2011). Analysis of the types and value measuring properties of the domestic Charolais herd. (in Hungarian) Doctoral dissertation, Szent István University, Gödöllő, Hungary, pp. 131. Retrieved from https://archive2020.szie.hu/file/tti/archivum/DomokosZoltan_ertekezes.pdf

Domokos, Z. & Tőzsér, J. (2014). Development of the Charolais cattle breed and types. (in Hungarian) Retrieved from http://www.charolais.hu/ujweb/index.php/hu/charolais-fajta/bemutatasa/18-charolais-fajta

Hámori, G. (2001). Characteristics of CHAID-based decision trees. (in Hungarian) *Statisztikai Szemle, 79(*8), 703-710. Retrieved from https://www.ksh.hu/statszemle_archive/2001/2001_08/2001_08_703.pdf

Horn, A. (ed.) (1976). *Cattle breeding*. (in Hungarian) Mezőgazdasági Kiadó, Budapest, Hungary, pp. 196-199.

Jánosa, A. (2011). *Data analysis using SPSS*. (in Hungarian) ComputerBooks, Budapest, Hungary, pp. 264-285. ISBN 9789636183684

Kass, G. (1980). An exploratory technique for investigating large quantities of categorical data. *Applied Statistics, 29*(2), 119-127. https://doi.org/10.2307/2986296

Magidson, J. & SPSS INC. (1993). *SPSS for Windows CHAID release 6.0.* SPSS Inc. Chicago, USA, Retrieved from https://search.worldcat.org/title/31617470

DOI: 10.17108/ActAgrOvar.2024.65.1.114

Sahin, O. (2023). Analysis of the effect of region, province and breed on somatic cell count in dairy cattle in Turkey by regression tree method. *Emirates Journal of Food and Agriculture, 35(*6), 588-597. https://doi.org/10.9755/ejfa.2023.v35.i6.3100

Tirink, C., Piwczynski, D., Kolenda, M., & Önder, H. (2023). Estimation of body weight based on biometric measurements by using random forest regression, support vector regression and CART algorithms. *Animals 13*(5), 795. https://doi.org/10.3390/ani13050798

Tőzsér J., Domokos Z., Alföldi L., Holló G., & Rusznák J. (2001). Body measurements of cows in a Charolais herd with different genotype. (in Hungarian) *Hungarian Journal of Animal Production, 50*(1), 15-22.

Yilmaz, I., Eyduran, E., & Kaygisiz, A. (2013). Determination of on-genetic factors influencing birth weight using regression tree methods in Brown-Swiss cattle. *Can J App Sci, 1*(3), 382-387. ISSN 1925-7430

Winston, W.L. (2003). Operations Research – Applications and Algorithms, 4th Ed., Thomson Press, pp. 641-645. Retrieved from https://itslearningakarmazyan.wordpress.com/wp-content/uploads/2015/09/operation-research-aplications-and-algorithms.pdf

DOI: 10.17108/ActAgrOvar.2024.65.1.114