International Journal of Agriculture and Environmental Research

ISSN: 2455-6939

Volume:04, Issue:06 "November-December 2018"

REGRESSION TREE ANALYSIS FOR PREDICTING RACE PERFORMANCE (SPEED) OF THOROUGHBRED RACEHORSES

[1*]Yıldırım F, [2]Yağanoğlu A. M., and [1]Yıldız A.

[1]Department of Animal Science, Faculty of Veterinary Medicine, University of Atatürk, Erzurum, Turkey

[2]Department of Animal Science, Faculty of Agriculture, University of Atatürk, Erzurum, Turkey

[*]Corresponding author

ABSTRACT

Today, the sport has become a more important use for horses. The aim of the present was to predict the race speed or performance of Thoroughbred racehorses. For this reason, using parameters of sex, race age, dam age, race track, race distance, city, horse age taken body measurement and body measurements, the aim is to use regression tree analysis to identify the most important predictor of race speed. The racing records (2445 races) of Thoroughbred racehorses (46 males and 28 females, total 74 horses) in this study were provided by the Turkey Jockey Club. As a result, putting into consideration the relative importance of race track in constructing regresyon tree, it could be inferred that Thoroughbred racehorses running turf track would be 'speedy'. In the races on the dirt track, the effect of cities on the horses' speed was found to be significant. Especially the high speed of horses in Istanbul has come into prominence among the provinces. In addition, cannon bone circumferences of 22.5 cm or smaller and chest circumferences greater than 174.5 cm in Thoroughbred horses were observed to positively affect the speed.

Keywords: horse, performance, regression tree analysis, speed, thoroughbred

INTRODUCTION

In the past, domesticated horses carried goods and hauled agricultural equipment, wheeled transport and sledges; when ridden, they were used for both transport and hunting. For this reason, power and endurance were the most desirable features for horses. Today, however, sport has become a more important use for horses, so that their care and nutrition has been aimed at improving their sporting performance. The process has changed equine body structure (Özbeyaz and Akçapınar, 2003).

ISSN: 2455-6939

Volume:04, Issue:06 "November-December 2018"

The oldest known running horse is Eohippus, a herbivorous animal of the Eocene period that emerged about 60 million years ago. Over the next several million years, these animals evolved to grow large; they occupied forest areas, eventually becoming steppe animals. After that, with the development of domestication, they adapted to new living conditions (Özbek, 2015).

When we examine the development of the Thoroughbred horse breed in England, we see that the most important founders of the new breed were Arabian stallions named the Darley Arabian, the Godolphin Arabian and the Byerly Turk. To combine these three horses' blood with the British native horse type took about fifty years. The Darley Arabian was the most influential of the three on the breeding of the Thoroughbred horse (Kısakürek, 2003).

In the Republic of Turkey, the Ministry of Agriculture and Forest has the authority to regulate horse racing and has assigned it to the Turkey Jockey Club (TJK). The TJK organizes separate races for Thoroughbred English and Thoroughbred Arabian horses (Paksoy and Ünal, 2010). There are meetings at nine official horseracing tracks throughout the country, some with turf and others with dirt surfaces (Özen and Gürcan, 2017).

The performance of a horse is closely related to the strength and flexibility of its body, as these confer the ability to move (Yıldırım, 2007). Body measurements in horses can be used to assess conformation or body harmony (Sadek et al., 2006). For example, in a study has been carried out on the evaluation of the body structure, as this is thought to be the most important contributor to the speed of the high-performance horse (Mawdsley et al., 1996).

Reportedly, the features to be looked for in an ideal racehorse are good body structure, speed and ambition. The racing career of a horse with unsuitable body structure is short (Stashak, 1987). The most important feature that we can use to reveal the performance and racing life fluency of a horse is its body structure, and this depends on the breed. Differences among breeds mean that only Thoroughbreds, all of the same breed, compete with each other (Arpacık, 1999).

Regression tree analysis is a nonparametric model that can explain the relationship between dependent continuous variable and independent continuous or categorical variables. Using an algorithm that minimizes variability, it repeatedly divides the data to determine homogeneous subgroups of independent variables (Zheng et al., 2009). The output of the regression tree analysis is a structure created according to the independent-variable-producing homogeneous nodes (Larsen and Speckman, 2004).

This paper aims to provide a way to predict the race speed or performance of Thoroughbred racehorses. Using parameters of sex, race age, dam age, race track, race distance, city, horse age taken body measurement and body measurements, the aim is to use regression tree analysis to identify the most important predictor of race speed as well as the optimum combination of the

ISSN: 2455-6939

Volume:04, Issue:06 "November-December 2018"

levels of predictors. It is hoped that this research will contribute to a better understanding of the effect of these factors on race performance.

MATERIALS AND METHODS

Animals: The racing records of Thoroughbred racehorses in this study were provided by the TJK and included official races that took place between May 2007 and December 2016. The dataset for this period was comprised of records from 2445 races. For every race, the information included each horse's sex, dam age, race track, race distance, race age and city. Detailed information for these variables is shown in Table 1.

Table 1: Data Parameters

Variable name	Category
Sex	Male, Female
Dam age (years)	4-7, 8-11, 12-15, 16-19, 20+
Race track	Dirt, Turf
Race distance (m)	1200, 1300, 1400, 1500, 1600, 1700, 1800, 1900, 2000, 2100, 2200
Horse race age (years)	3, 4, 5, 6+
Race city	Adana, Ankara, Bursa, Diyarbakir, Elazığ, Istanbul, Izmir, Sanliurfa

Body measurement: Body measurements were taken from 74 Thoroughbred racehorses (46 males and 28 females) registered by the TJK stud farms in Adana, Sanlıurfa, Bursa and Ankara. The age groups used for each set of body measurements were 3, 4, 5 and 6+ years. Thirteen body parts on each horse were measured, with the horse in a normal standing position on a concrete or other flat surface. Withers height (WH), croup height (CH) and chest depth (CD) were measured with a measuring stick (Hauptner®); croup length (CL) and pectoral chest width (PCW) were measured with a measurement gauge (Elmark®); and back length (BAL), body length (BL), chest circumference (CC), neck circumference (NC), head length (HL), cannon bone circumference (CBC), carpal joint circumference (CJC) and hock circumference (HC) were measured with a measuring strip (Yıldırım and Yıldız, 2013).

Statistical analysis: This study analysed the effects of variables thought to affect a horse's race speed (measured in metres per second). These variables are: sex, race age, dam age, race track, race distance, city, horse age taken body measurement and body measurements. Regression tree analysis was performed using the SPSS statistical package program (SPSS 20.0 for Windows).

ISSN: 2455-6939

Volume:04, Issue:06 "November-December 2018"

We give a brief overview here of regression tree methodology, assuming that the reader is familiar with the basic principle. In the classic Classification Analysis and Regression Tree (CART) program of Breiman et al. (1984), a greedy search algorithm is used to construct a binary tree in the independent variables. The goal is to produce nodes as homogeneous as possible with respect to the dependent variable. Consider the multiple regression problem $y_i = f(x_{i1}\ldots\ldots x_{ip}) + \varepsilon_i$, $i = 1,\ldots\ldots,n$, where f is unknown and not easily parameterized. x_{ij} are known independent variables, and ε_i are random error terms with zero means. A node N is a subset of the indices $\{1,\ldots.., n\}$. The deviance of a node N is defined as:

$$D(N)=\sum_{i\epsilon N}\{ y_i-\bar{y}(N)\}^2$$

where $\bar{y}$ is the mean of observations in T node and y_i is the value of the ith observation (Larsen and Speckman, 2004; Questier et al., 2004).

RESULTS

Table 2 shows the significance levels of independent variables included (or not included) in the regression tree structure affecting horse speed. The improvement-based importance value of each variable indicates the surrogate for the primary splitting variable. Surrogate variables result in similar splits in both nodes, as does the primary variable. The importance value indicates the influence on the dependent variable of variables whose effect is masked by other variables on the regression tree. Relative importance values are calculated as a proportion of the importance of each value to the first importance value. Thus, relative importance value indicates the influence of each variable on the dependent variable when each variable is used to replace the most important variable.

Table 2: The significance of the independent variables affecting horse speed

Independent Variable	Importance	Relative Importance (%)
Race track	0.134	100.00
Race city	0.05	37.20
Sex	0.021	15.60
Chest dept (cm)	0.019	13.80
Croup height (cm)	0.015	11.40
Hock circumference (cm)	0.014	10.40
Back length (cm)	0.013	9.60
Neck circumference (cm)	0.012	9.10
Cannon bone circumference (cm)	0.012	8.90
Chest circumference (cm)	0.012	8.80

ISSN: 2455-6939

Volume:04, Issue:06 "November-December 2018"

Withers height (cm)	0.011	8.10
Pectoral chest width (cm)	0.009	7.00
Head length (cm)	0.007	5.50
Carpal joint circumference (cm)	0.007	5.10
Horse race age (years)	0.004	3.30
Croup length (cm)	0.003	2.50
Body length (cm)	0.001	0.80
Race Distance (m)	0.001	0.60
Dam age (years)	0	0.10
Horse age taken body measurement (years)	0	0.10

Figure 1 shows the regression tree diagram of the factors expected to affect the speed of Thoroughbred racehorses. In the figure, descriptive values of the speed are given in the main node of the regression tree (Figure 1). Average speed ± Standard Deviation in Thoroughbred racehorses was found to be 15.393 ± 0.771 m/s. The primary node was divided into two nodes by the race track variable, indicating that the race track variable has the most influence on speed. It was found that the race track node 1 (dirt track) had a total of 1775; the average speed of these horses was 15.167 ± 0.702 m/s. The turf track (terminal node 2) within the enterprise was 27.4% and the speed was 15.990 ± 0.608 m/s. Node 1 was further split into two child nodes (3 and 4) according to the city. The regression tree diagram indicates that the city is the second variable affecting the speed of the horses. It was found that the average speed in Bursa, Istanbul, Ankara, Izmir and Adana (node 3) was 15.300 ± 0.594 m/s, and in Elaziğ, Sanıurfa and Diyarbakir (terminal node 4) it was 14.824 ± 0.834 m/s. Node 3 was further split into two nodes (5 and 6) by the city variable. In terminal node 5, the average speed of horses in Istanbul was 15.713 ± 0.567 m/s. In node 6, the average speed of horses in Bursa, Ankara, Izmir and Adana was 15.256 ± 0.580 m/s. Node 6 was split into two child nodes (7 and 8) according to cannon bone circumference. It was found that in node 7, 1068 horses with cannon bone circumferences less than or equal to 22.5 cm had an average speed of 15.287±0.578 m/s, while horses with cannon bone circumferences greater than 22.5 cm (terminal node 8) had an average speed of 14.900 ± 0.475 m/s. Node 7 was split into two child nodes (9 and 10) according to chest circumference. It was found that in node 9, 538 horses with chest circumferences less than or equal to 174.5 cm had an average speed of 15.191±0.558 m/s, while those with chest circumferences greater than 174.5 cm (node 10) had an average speed of 15.383 ± 0.582 m/s. In addition, child nodes 9 and 10 could not be split further and are thus terminal nodes.

ISSN: 2455-6939

Volume:04, Issue:06 "November-December 2018"

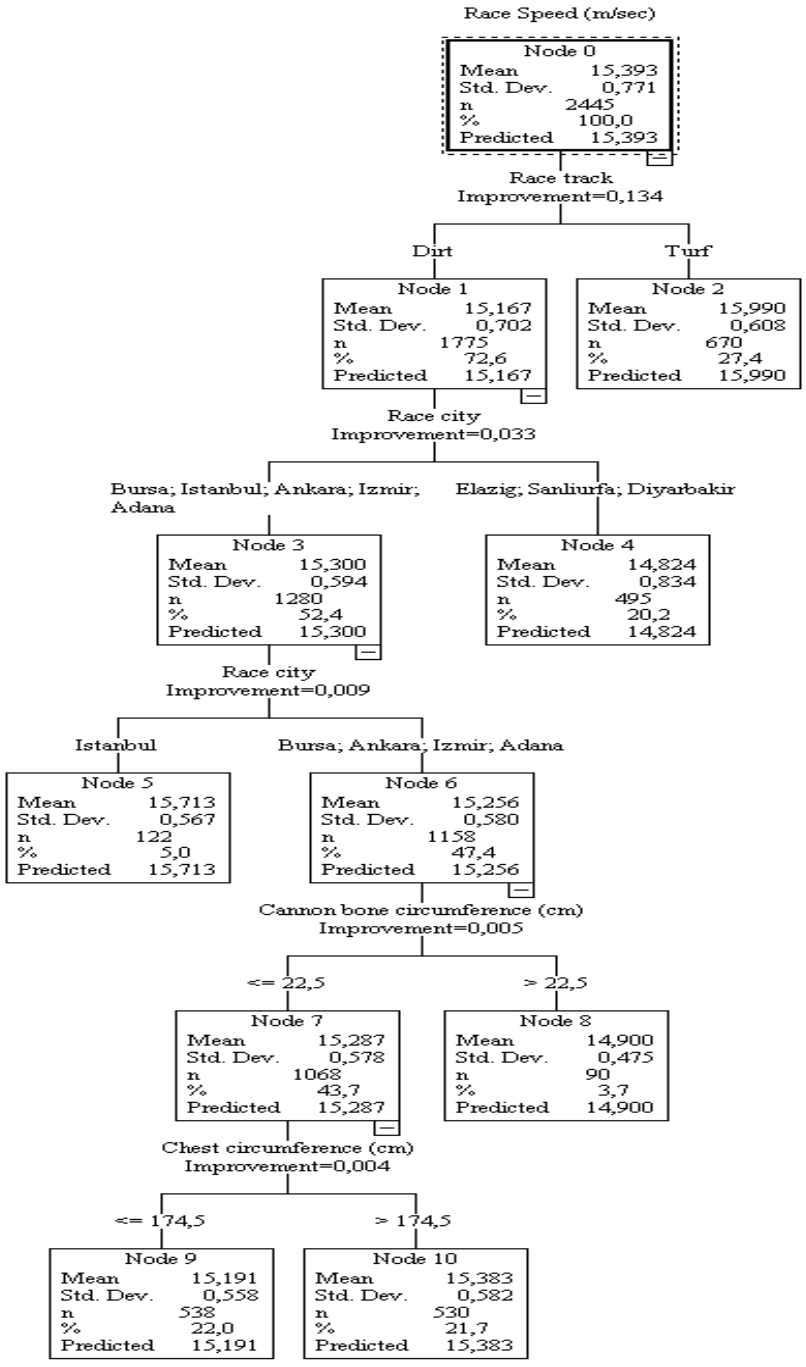

Figure 1: Regression tree analysis of the factors affecting speed of Thoroughbred racehorses

ISSN: 2455-6939

Volume:04, Issue:06 "November-December 2018"

DISCUSSION

The average race speed (node 0, 15.393 m/s) obtained in this study was smaller than the average race speed (16.3 m/s) of Thoroughbreds racing in Western Australia (Crispe et al., 2017), but was similar to the race speed range (14.73–15.76 m/s) of Thoroughbreds racing in Turkey (Paksoy and Ünal, 2010).

It could be inferred from the regression tree that of all the indices (sex, race age, dam age, race track, race distance, city, horse age taken body measurement and body measurements) included as prediction variables, the race track contributed most significantly to the estimation of the horse speed.

This research has found a direct relationship between the race track and any given Thoroughbred horse's race speed; this was true in all cases. Paksoy et al. (2018) compared Thoroughbred horses' race speeds on turf and dirt surfaces, obtaining results that corroborated the findings of this research. They found that the speeds varied with the track type, with race speed on dirt tracks being generally lower than on the turf. However, the amount of moisture in either track type also affects the speed of the horses. Paksoy et al. (2018) concluded that the horses' pedigrees and performance records should be considered when deciding which horses are likely be more successful than others.

When the effect of the city on the other factors affecting the race speed was examined, horses running at the cities of Bursa, Istanbul, Ankara, Izmir and Adana were observed to be faster than those running at Elazig, Sanliurfa and Diyarbakir. Yıldırım (2019), who examined the race performance of Arabian horses in the cities covered by this study, stated that the altitude of the venue affected the horses' racing speed, noting that, in general, slow speed correlated with high altitude. This study's results paralleled those of Yıldırım (2019), showing that the speed or performance of Thoroughbred horses is affected by the altitude of the racing location. However, in node 3 (Bursa, Istanbul, Ankara, Izmir, Adana), the horses' speed in Istanbul city was found to be higher than it was in other cities. This case indicates that maintenance, feeding and riders, among other factors, may affect the speed of the race.

Even among horses of the same breed, differences in body structure between individuals may have some effect on race speed or performance. The various parts of the body (head, neck, chest, back, withers, croup and abdomen) and their harmonies have an effect on horses' performance (Paksoy and Ünal, 2010). In horses, the most commonly recorded body measurements are WH, CH, CD, CL, PCW, BAL, BL, CC and CBC. In this study on Thoroughbred horses, the cannon bone circumference (CBC) and chest circumference (CC) were included in the regression tree model.

ISSN: 2455-6939

Volume:04, Issue:06 "November-December 2018"

In the mid 1900s, Yarkın (1962) emphasized that a short, wide cannon bone was a desirable feature because it would increase the horse's endurance. Nowadays, however, good speed and performance is expected as well as endurance. In this study, it was observed that racehorses with a CBC equal to or less than 22.5 cm were faster and performed better than those with a greater CBC. We can therefore say that the cannon bone circumferences favoured in the antiquity did not have a positive effect on the racing speed of today's Thoroughbred horses.

The health and efficiency of the animals depend heavily on the chest circumference, together with the condition and size of the organs within the chest (Akçapınar and Özbeyaz, 1999). This study confirmed that an increase in chest circumference positively affects race speed. Horses with a chest circumference greater than 174.5 cm showed better racing speeds or performances.

Some researchers have shown that the length of the long bones, strongly related to the height at the withers, does not increase between the ages of two and three years (Andersson and Mcılwraith, 2004). This study, examining the effect of body measurements on the racing speed, evaluated only those horses aged three years and above, because it is accepted that the maturation rate of Thoroughbred horses is greatly reduced once they have reached three years. The results show that, after that age, the effect of age on race speed is minimal. Body measurements taken at or after the age of three are therefore appropriate for evaluating race performance or speed.

CONCLUSION

The regression tree analysis has demonstrated the practical possibility of combining sex, race age, dam age, race track, race distance, city, horse age, horse age taken body measurement and body measurement in predicting the race speed or performance of Thoroughbred racehorses. Considering the relative importance of the race track in constructing the regression tree, it could be inferred that Thoroughbred racehorses would be 'speedy' when running on a turf track. In dirt-track racing, the effect of the city on the horses' speeds was found to be significant, and the high speed of horses in Istanbul has come into particular prominence among the provinces. In addition, cannon bone circumferences of 22.5 cm or smaller and chest circumferences greater than 174.5 cm were observed to have positive effects on the speed of Thoroughbred horses.

Acknowledgments

The authors would like to thank the employees of the Turkish Jockey Club for their collaboration in providing access to data used in this study.

REFERENCES

1. Akçapınar H, Özbeyaz C. 1999. Hayvan Yetiştiriciliği Temel Bilgileri. Kariyer Matbaacılık, Ankara.
2. Andersson TM, Mcılwraith CW. 2004. Longitudunal development of equine conformation from weanling to age 3 years in the Thoroughbred. Equine Veterinary Journal, 36(7), 563–570.
3. Arpacık R. 1999. At Yetiştiriciliği. 3. Baskı. Şahin Matbaası. Ankara.
4. Breiman L, Friedman JH, Olshen RA, Stone CI. 1984. Classification and regression trees. Belmont, Calif.: Wadsworth.
5. Crispe EJ, Lester GD, Secombe CJ, Perera DI. 2017. The association between exercise-induced pulmonary haemorrhage and race-day performance in Thoroughbred racehorses. Equine veterinary journal, 49(5), 584-589.
6. Kısakürek NF. 2003. At'a Senfoni. Türkiye Jokey Kulübü Yayınları. 71-78.
7. Larsen DR, Speckman PL. 2004. Multivariate Regression Trees for Analysis of Abundance Data. Biometrics, 60, 543-549.
8. Mawdsley A, Kelly EP, Smith FH, Brophy PO. 1996. Linear assessment of the Thoroughbred horse: A approach to conformation evalution. Equine Veterinary Journal, 28(6), 461–467.
9. Özbek S. 2015. Benim Atım. Türkiye Jokey Kulübü Yayınları.
10. Özbeyaz C, Akçapınar H. 2003. At yetiştiriciliği ders notları. Ankara Üniversitesi Veteriner Fakültesi, Ankara.
11. Özen D, Gürcan İS. 2017. Factors that affect whether Arabian horses have earnings during their first year of racing. Turk J Vet Anim Sci, 41(4), 460-463.
12. Questier F, Put R, Coomans D, Walczak B, Vander Heyden Y. 2004. The use of CART and multivariate regression trees for supervised and unsupervised feature selection. Chemometrics and Intelligent Laboratory Systems, 76, 45-54.
13. Paksoy Y, Ünal N. 2010. Atlarda Yarış Performansını Etkileyen Faktörler. Lalahan Hay Araşt Enst Derg, 50(2), 91-101.
14. Paksoy Y, Ünal N, Polat M, Tekin M, Özbeyaz C. 2018. Investigation of effect of hoof size on racing performance in Arabian and Thoroughbred horses. Lalahan Hayvancılık Araştırma Enstitüsü Dergisi, 58(1), 22-33.
15. Sadek MH, Al-Aboud AZ, Ashmawy AA. 2006. Factor analysis of body measurements in Arabian horses. Journal of Animal Breeding Genetics, 123, 369–377.
16. Stashak TS. 1987. The Relationship Between Conformation and Lameness. In: Adam's Lameness in Horses, 4th edition. Ed: TS Stashak. Lea and Febiger, Philadelpia. P 71.
17. Yarkın İ. 1962. Atçılık. Ankara Üniv. Zir. Fak. Yayınları, Yayın No:40.

ISSN: 2455-6939

Volume:04, Issue:06 "November-December 2018"

18. Yıldırım F. 2019. The Effect of Some Environmental Factors To Race Performance of Purebred Arabian Horses. Kocatepe Veteriner Dergisi, (in press). DOI: 10.30607/kvj.450350
19. Yıldırım F, Yıldız A. 2013. Body measurements in the javelin horses. Kafkas Univ Vet Fak Derg, 19(4), 693-698.
20. Yıldırım İG. 2007. Atlarda genel vücut yapısının morfometrik yöntemlerle incelenmesi (Doctoral dissertation, Adnan Menderes Üniversitesi).
21. Zheng H, Chen L, Han X, Zhao X, Ma Y. 2009. Classification and regression tree (CART) for analysis of soybean yield variability among fields in Northeast China: The importance of phosphorus application rates under drought conditions. Agriculture, Ecosystems & Environment, 132, 98-105.

