IJVAR International Journal of Veterinary and Animal Research
Uluslararası Veteriner ve Hayvan Araştırmaları Dergisi
E-ISSN:2651-3609 1(2): 47-50, 2018

Possibilities Of Estimating Body Weight From Different Body Measurements In Hair Goat Using Different Regression Models

Yasin DEMIRASLAN[1*] Iftar GURBUZ[1] Omer Gurkan DILEK[1] Sukru OZUDOGRU[2] Ozcan OZGEL[1]
[1]Department of Anatomy, Faculty of Veterinary Medicine, Mehmet Akif Ersoy University, Burdur, TURKEY

[2] Department of Archaeology, Faculty of Science and Literature, Mehmet Akif Ersoy University, Burdur, TURKEY

*Corresponding Author
E-mail:yasindemiraslan@hotmail.com

Abstract

In the study, the data obtained to describe the body characteristics of the Hairpin were utilized in the businesses that were registered with Karaman Province Breeding Sheep Goat Breeders Association. Body weights of 130 goats, 2, 3, 4, 5, 6 and 7 years old and 50 goats, 2, 3 and 4 years old, selected by simple random sampling method were used in the data of total 900. In the study, Pearson correlation coefficient for variables providing parametric test prerequisites, and Spearman correlation analysis for variables not providing parametric test prerequisites. In the regression analysis, "live weight" dependent variable and other variables were determined as independent variables and parametric and nonparametric regression methods were applied. Univariate and multivariable regression models were applied for the whole data set. When all analyzes are evaluated, univariate regression models give lower determination coefficients (R2) than multivariate models. In this case, it has been deemed appropriate to use a multivariate regression model instead of a univariate model in order to make a correct prediction.However, in practice, univariate Quadratic or Cubic regression methods can be used for researchers.

Keywords: Regression, Semi-parametric regression models, Parametric regression models, Live weight in the goats, Body measurements

INTRODUCTION

Regression analysis assumes that when the mean relation between the dependent variable and the independent variable is expressed by a mathematical function, the independent variable and the dependent variable are in a linear relationship.

Regression models are regression models known as parametric, nonparametric and semi-parametric regression methods.

All of the approaches available for the semi-parametric regression model are based on different non-parametric regression methods. Semi-parametric regression models summarize complex data sets in a way that we can understand and maintain important properties while ignoring the insignificant details of the data in practice, thus allowing robust decisions to be made [1].

Semi-parametric regression method is widely used in the analysis of time-dependent data. Generally, longitudinal data obtained from experiments in the fields of agriculture, medicine and biostatistics are measured with a continuous scale depending on the time, and measurements taken at different times from the same trial unit (individual) take different values. But the recipients are related to each other. This is the result of applying multiple behaviors to the same test units to follow each other [2].

In the majority of longitudinal studies, the effects of time and continuous independent variables on the resulting outcome variance are included in the model. Correlation (autocorrelation) between error variables occurs when more than one observation is made on the same individual depending on location and time. In such cases, some assumptions do not apply. Therefore, making time-related assessments is a common problem for parametric methods. Non-parametric methods can be used in such cases. However, when nonparametric methods are used to analyze the number of independent variables, it is difficult to make analyzes and to interpret the graphs. As an alternative method, semi-parametric models can be used. In semiparametric models, the effects of chance and time are affected by nonparametric methods, while the effects of continuous independent variables are included by methods that are parametric.

The semi-parametric regression model is also called the "partial linear model" by the fact that it consists of a combination of parametric and non-parametric regression functions. In the study, the live body weight was estimated from different body measurements in the hair follicle by the multivariate, univariate parametric and nonparametric regression methods.

MATERIALS AND METHODS

SIn regression analysis, there are two types of linearity in variables and coefficients (linearity in parameters). The state of linearity in variables means that the value of each variable in the model is one; indicates a linear functional relationship between dependent and independent variables. Similarly, in coefficients, linearity is the exponent of all coefficient values in the model and the existence of a linear functional relationship between the dependent variable and the coefficient values.

$$Y_i = \beta_0 + \beta_1 X_i + e_i \tag{1}$$

An example of a model is that both the coefficients and the variables are linear.

$$Y_i = \beta_0 + \beta_1 X^2{}_i + e_i \tag{2}$$

The coefficients are also linear, but the variables are examples of nonlinear models.

$$Y_i = \beta_0 + \sqrt{\beta_1} X_i + e_i \tag{3}$$

Variables are linear, while coefficients are examples of nonlinear models.

Simple Linear Regression Model

The regression model examines the causality relationship between a single independent variable and a dependent variable.

$$Y_i = \beta_0 + \beta_1 X_i + e_i \qquad (4)$$

Multiple Regression Model

Models developed for multiple regression analysis resemble simple linear regression models, with the exception of more terms, and can be used to examine straightforward, more complex relationships. For example, suppose that the average time E (y) needed to fulfill the data-processing task increases as the use of computers increases and we think that the relationship is curve-linear. $E(y) = \beta_0 + \beta_1 X_1$

To model the deterministic component, the following quadratic model can be used instead of the straight-line model.

$$E(y) = \beta_0 + \beta_1 X_1 + \beta_2 X_1^2 \qquad (5)$$

For example, the first-order model

$$E(y) = \beta_0 + \beta_1 X_1 + \beta_2 X_2 \qquad (6)$$

(x1, x2) -plane. For our example (and for many real-life applications), we expect a slope on the response surface and use a second-order model to model the relationship.

$$E(y) = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_1 X_2 + \beta_4 X_1^2 + \beta_5 X_2^2 \qquad (7)$$

All the models written up to now are called generic linear models, because E (y) is a linear function of unknown parameters. The following model is not linear.

$$E(y) = \beta_0 e^{-\beta_1 x} \qquad (8)$$

Because E (y) is not a linear function of unknown model parameters.

Semi-parametric Regression Models

Semi-parametric regression models are models in which the dependent variable can be parameterized in relation to some explanatory variables, but not easily related to some other explanatory variable or variables. In the semi-parametric model, linear parametric components form the parametric part of the model whereas both parametric and non-linear components form the non-parametric part of the model. This model is a special case of additive regression models [3], which allows easier interpretation of the effect of each variable and generalizes standard regression methods. In addition, the semi-parametric model is a model in which the dependent variable is linear with some explanatory variables but not linear with other specific independent variables.

Parametric Methods

Linear: $Y = b_o + b_1 X$ (9)

Inverse: $Y = b_0 + (b_1 / t)$ (10)

Quadratic: $Y = b_o + b_1 X + b_{11} X^2$ (11)

Cubic: $Y = b_o + b_1 X + b_{11} X^2 + b_{111} X^3$ (12)

Semi-Parametric Methods

Logarithmic: $Y = b_0 + (b_1 * \ln(t))$ (13)

Power: $Y = b_o + b_1 X$ veya $\ln(Y) = \ln(b_0) + (b_1 * \ln(t))$ (14)

Compound: $Y = b_0 * (b_{11}^2)$ veya $\ln(Y) = \ln(b_0) + ((b_1) * \ln(t))$ (15)

S-curve: $Y = e^{(b_0 + (b_1/t))}$ veya $\ln(Y) = b_0 + (b_1/t)$ (16)

Growth: $Y = e^{(b0 + (b1 * t))}$ veya $\ln(Y) = b_0 + (b_1 * t)$ (17)

Exponential: $Y = b_0 * (e^{(b1 * t)})$ veya $\ln(Y) = \ln(b_0) + (b_1 * t)$ (18)

Y= dependent variable

b_0= Regression equation constant

b_1= Regression coefficient

t= numeric value of the independent variable

Multivariate Regression Models

Variable Selection Methods

(ForwardSelection)

(BackwardElimination)

(Stepwise Regression)

$$E(y) = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \ldots + \beta_k X_k \qquad (19)$$

[4,5,6]

In the survey, the data obtained for the purpose of describing the body characteristics of the Hairpin were utilized within the scope of "Project for the development of subspecies of the hairpin race", "Project code: Tagem / Kıl2013-02", in the enterprises that have registered the Karpil breed sheep goat breeders association in Karaman province. The body measurements of 130 goat selected by simple random sampling method of 2, 3, 4, 5, 6 and 7 aged females were used in this study and a total of 50 teens data selected by simple random sampling method of 2, 3 and 4 elderly monopolies were used for monopolies.

The live weights of the goats and body measurements were taken at the end of the forties in June. Body measurements were made in the cage or on the flat surface of the cave.

The body measurements measured by goats in 2012, the measurements made and their anatomical definitions are given below.

Height at withers (CY)

Height at rump (SY)

Body length(VU)

Rump Width(SG)

Chest width (GG)

Chest depth(GD)

Chest girth (GÇ)

Pearl Circle (İÇ)

The data to be used in the study were randomly selected from the general data with the MINITAB program. Statistical package program Syntax Function SPPS 20 (IBM Corp. Released 2011. IBM SPSS Statistics for Windows, Version 20.0, Armonk, NY: IBM Corp.) was used to evaluate the data. The level of significance is shown as α = 0,05.

FINDINGS

Figure 1. Univariate parametric and semi-parametric regression model graphs

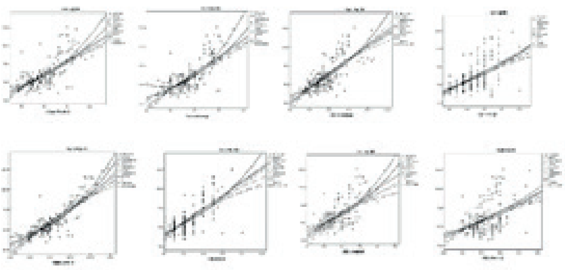

Table 1. Results of univariate parametric and semi-parametric regression models

	Methods	Summary Model					Estimation of parameters				
		R^2, %	F	SD	SD	p	Sabit	b1	b2	b3	
Height at withers	Quadratic	49,8	87,6	2	177	0,001	246,8	-6,4	0,051		$Y = 454,5 - 12,1X + 0,09X^2$
Height at rump	Quadratic	60,4	135,2	2	177	0,001	454,5	-12,1	0,09		$Y = 454,5 - 12,1X + 0,09X^2$
Body length	Linear	54	209	1	178	0,001	-89,88	1,93			$Y = -31,78 + 5,001X$
Rump Width	Cubic	31,3	40,2	2	177	0,001	4,79	0	0,22	-0,003	$Y = 37,28 + 0,001X + 0,028X^2 - 0,002X^3$
Chest width	Cubic	46,6	76,8	2	177	0,001	37,28	0,001	0,028	0,002	$Y = 37,28 + 0,001X + 0,028X^2 - 0,002X^3$
Chest depth	Logarith-mic	24,9	50	1	178	0,001	-247,5	86,06			$Y = -247,50 + (86,06 * \ln(t))$
Chest girth	Quadratic	79,8	349,2	2	177	0,001	235,56	-5,48	0,039		$Y = 235,56 - 5,48X + 0,039X^2$
Pearl Circle	Linear	51,1	186,1	1	178	0,001	-30,6	8,7			$Y = -30,6 + 8,7X$

Table 2. Result of multivariate regression model (Stepwise method)

Model		Coefficients		t	p	95,0% Confidence Interval		R^2, %	P
B		S. Error			Alt Sınır	Üst Sınır			
1	(Constant)	-106,2	7,19	-14,8	0,001	-120,4	-92,0	75,3	0,001
	Chest girth	1,8	0,08	23,3	0,001	1,7	2,0		
2	(Constant)	-122,7	7,58	-16,2	0,001	-137,7	-107,7	78,2	0,001
	Chest girth	1,5	0,11	14,0	0,001	1,3	1,7		
	Body length	0,6	0,13	4,8	0,001	0,4	0,9		
3	(Constant)	-118,4	7,45	-15,9	0,001	-133,1	-103,7	79,7	0,001
	Chest girth	1,3	0,11	11,5	0,001	1,1	1,5		
	Body length	0,5	0,13	3,7	0,001	0,2	0,8		
	Pearl Circle	2,1	0,60	3,5	0,001	0,9	3,3		
4	(Constant)	-114,0	7,69	-14,8	0,001	-129,1	-98,8	80,1	0,001
	Chest girth	1,2	0,13	9,3	0,001	0,9	1,4		
	Body length	0,5	0,13	3,6	0,001	0,2	0,7		
	Pearl Circle	2,0	0,59	3,5	0,001	0,9	3,2		
	Chest width	0,5	0,23	2,1	0,041	0,02	0,9		

When the estimation equations for univariate methods are examined, Quadratic or Cubic models give higher R2 value, unlike the use of continuous linear models (Table 1 and Figure 1).

As a result of the multivariable regression methods, it is possible to estimate body weight by 80% with the regression equations generated by independent variables of Chest Environment, Body Length, Hip Circumference and Chest Width. As a result of univariate regression methods, Quadratic or Cubic models predict body weight by about 75% with independent breast circumference variation. Multivariate regression methods result in an increase of 5% when the Body Length, Thigh Circumference, and Chest Width arguments are added (Table 2) [7,8].

DISCUSSION and SUGGESTIONS

Some criteria are relevant to determine which statistics are applicable to the data obtained in a study. Analyzing the research with appropriate statistical methods also improves the reliability of the research and provides a consistent interpretation of the results. For this reason, variable structures, measurement scales, and consistency of assumptions are important considerations in statistical studies.

Using inappropriate regression methods can lead to incorrect and misleading results. The relationship between variables must be examined with functional regression models. The regression model that needs to be used differs according to the structure of the data, and using the wrong model can lead to incorrect results. In this case, it is suggested to establish the most meaningful model suitable for data structure.

In the study, differences in the mean of the best model were observed among the results of the different body regimens included in the model as the univariate independent variable versus the live weight dependent variable, in the different regression models applied. In all body dimensions, all linear and non-linear models were found to give statistically significant results. It has been seen that most of the body measurements give more favorable results in the sense of both R2 and Cubic models. Only in the chest depth variable the logarithmic model gave the highest R2 value. It is understood that the Quadratic or Cubic model can be preferred to the Linear model because all variables except this give the equal R2 value of the body length and width of the rider which can be preferred to the Quadratic model.

It is predicted that multivariable regression equations generated by independent variables of Chest Environment, Body Length, Thigh Circumference and Chest Width can be prefered as a result of multivariate stepwise and best subset regression analyzes in the study.

REFERENCES

1- Ruppert, D., Wand, M.P., Carroll, R.J., 2003. Semiparametric regression. Cambridge University Press.

2- Lee, T.C.M., Solo, V., 1999. Bandwith selection for local linear regression: a simulation study. Comput. Statist., 14, 515-532.

3- Härdle, W., Muller, M., Sperlıch, S., Werwatz, A., 2004. Nonparametric and semiparametric models. Springer, New York.

4- Farebrother, R.W., 1976. Further results on the mean square error of ridge regression. J.R Stat .Soc. B, 38:248-250.

5- Rao, C.R., Toutenburg, H., 1995. Linear models: least squares and alternatives. Springer Verlag, New York.

6- Robınson, M.P., 1988. Root-N-Consistent semiparametric regression Econometrica, 56(4): 931-954.

7- Tabakan, G., Akdenız, F., 2008. Difference based ridge estimator of parameters in partial linear model. Statistical Papers, DOI: 10.1007/s00362- 008-0192-6 (SCI Expanded).

8- Ünalan, A., Ceyhan A., 2017. Kilis Keçilerinin Canlı Ağırlık ve Bazı Vücut Ölçüleri Üzerinde Cinsiyet Etkisinin Belirlenmesi, Harran Tarım ve Gıda Bilimleri Dergisi, 21(2): 219-226.