RESEARCH ARTICLE
Journal of Biometry Studies (2021) 1(1): 23-25
DOI: 10.29329/JofBS.2021.348.05
www.biometrystudies.com

Journal of Biometry Studies

Parameter estimation of factors affecting milk yield in a multicollinearity by Bayesian Regression method

Aycan Mutlu YAĞANOĞLU[1,*]

[1]Ataturk University, Faculty of Agriculture, Department of Animal Science, Erzurum/TURKEY

[*]Corresponding author: myagan@atauni.edu.tr

Received: 15/06/2021, Accepted: 26/06/2021

Abstract

The regression analysis determines the relationship model between dependent and independent variables. In this study, the body weight, milking time, milk yield and environmental factors obtained from 42 dairy cattle were used for internal and external temperatures. In this study, the Bayesian Regression method was used to estimate milk yield parameters in case of multicollinearity. According to the results, it was seen that Bayesian method can be applied successfully in the field of animal husbandry. It is thought that the use of this study for dairy cattle in other agricultural areas will be useful for better evaluation of the data obtained.

Keywords: Bayesian regression, Multicollinearity, Milk yield

Please cite this article as follows:
Yağanoğlu, A. M. (2021). Parameter Estimation of Factors Affecting Milk Yield in a Multicollinearity by Bayesian Regression Method. *Journal of Biometry Studies*, *1*(1), 23-25. https://doi.org/10.29329/JofBS.2021.348.05

1. Introduction

The main purpose of cattle breeding is to achieve the highest yield in the most economical way as in other species. The level of yield is determined by two main factors, the animal's genetic structure and environmental conditions. These two basic elements need to be considered together to maximize the manufactured product (Jensen, 2001).

Friesian breed in the numerical sense in that culture comes the first race of cattle breed in Turkey. However, as in other breeds, the yield levels of cattle breeds are largely affected by environmental conditions as well as their genetic structure. As a matter of fact, different levels of efficiency are obtained in different regions of our country and under different operating conditions. This shows to some extent the level of regions and businesses in cattle breeding (Erdem et al., 2007).

Recently, 305-day milk yield (P305) was replaced by test-day milk yield (Ferreira et al., 2002), with the latter approach showing several advantages: a) it permits the removal of environmental variation in phenotypic data on milk yield, since test-day milk yield considers the specific environmental effects for each production record, which is not possible when P305 data are used (Visscher & Goddard, 1995); b) it grants a more accurate evaluation of cows, due to the use of a larger number of records per cow, as compared to the same records fitted to P305 (Rekaya et al., 1999); c) it is not affected by the accuracy of the different prediction methods for P305 (Rekaya et al., 1999), because it permits the use of part lactation information, without the need for adjusted factors and/or lactation prediction; d) it facilitates the genetic evaluation of lactation persistency (Jensen, 2001); e) it permits a more accurate estimation of the genetic and permanent environmental effects that affect milk yield.

Regression analysis is a statistical analysis technique that characterizes the relationship between two or more variables with a cause-effect relationship with a mathematical model called regression model to make estimations or predictions about that subject. After the regression model is fitted, checking whether the model is sufficient is the most important part of regression analysis. It is necessary to ensure that the fitted model approximates the correct model sufficiently and to check whether the least-squares regression analysis meets all its assumptions.

If the regression model does not fit adequately, it will give weak or misleading results. In addition to the variance analysis and R^2 which are generally used to determine the adequacy of the model in regression analysis, other tests can provide more explanatory information but are not used much in practice (Draper & Smith, 1981).

2. Material and Methods

The milk yield and records of 42 head of Simmental, Holstein and eastern Anatolian red cows, which were grown in 2 private agricultural enterprises within the borders of Erzurum, between January and March 2020 constituted the research material. In the research, internal temperature (IT), external temperature (ET), live weight (LW) and milking period (MP) were emphasized as milk yield characteristics.

2.1. Bayesian method

Bayes' theorem has recently been used in decision making techniques. Since the aim is to minimize the risk of making wrong decisions as much as possible, it would make sense to participate in the decision-making process in personal experience and knowledge (Vila et al., 2000).

Bayesian method is;

$$P(B\backslash A) = \frac{P(BA)}{P(A)} \quad (1)$$

In the same way, we use the Bayes theorem when finding conditional probability density functions. Here for probability density functions for θ and y (Zellner, 1971),

$$f(\theta|y) = \frac{f\big(\theta\, f(y \mid \theta)\big)}{f\ (\,y\,)} \quad (2)$$

Using the Bayesian estimator;

$$\hat{\beta}_b = (\sigma^{-2}(X'X)\beta + \sigma^{-2}A_O\beta_O)(\sigma^{-2}(X'X) + \sigma^{-2}A_O)^{-1} \quad (3)$$

The analysis to be made depending on various assumptions about the average of the preliminary distribution to be defined for the parameters also varies (Leamer, 1973).

2.2. Least square method

Generally a linear regression model;

$$Y = X\beta + \varepsilon \quad (4)$$

Here; Y: (nx1) dimensional chance variable vector, X: (nxp) dimensional known coefficient matrix, β: (nx1) size unknown parameter vector, ε: (nx1) size random variable vector, its mean is zero ($E(\varepsilon)=0$) and its variance ($var(\varepsilon)=\sigma^2 I$) is constant (Atkinson, 2000), n: number of observations, p: number of parameters.

Collinearity; coexistence term can be defined as the linear dependence of the columns of the matrix X. The problem of interchangeability between independent variables, a) incorrect data collection method, b) population or model constraints can be caused, c) errors in defining the model, and d) errors in model selection. This problem in the data may cause the variance and covariance values of the least-squares estimators of the regression coefficients to be large and consequently, the interpretations based on the regression model to be erroneous (Montgomery & Peck, 1992).

Graphical method and DW d-statistics (1950, 1951) are most commonly used to determine the autocorrelation problem in linear regression analysis. Other important methods include Durbin-h (Durbin, 1970), Breusch-Godfrey LM (Breusch , 1978), Ljung-Box Q (Ljung & Box, 1978), Rao F (Rao, 1973) and runs tests for generally small volumes of samples.

3. Results

Since the VIF values are greater than 10, there is multicollinearity in the data (Table 1). In the study, the values obtained by the Least Squares (LS) and Bayes Approach (BA) of the milk yield model and the Sum of Error Squares of the parameters are shown in Table 2. Since the table is analyzed, it is observed that there are differences according to the methods in terms of parameter values and error square sum values. The "a" parameter, which shows the highest asymptote value that the varieties take values between 5364,142 and 5311,575 in its method, while in the Bayesian method it has values between 5360,244 and 5311,575. When the asymptotic values were analyzed, the highest value was observed in the same type with 5364,142 values in the LS method and 5360,244 in the Bayesian method. This variety was followed for Simmental by 5313,738 in LS method and 5311,575 in Bayesian method. The highest asymptotic value was observed in Holstein with the values of 5364,142 in the LS method and 5360,244 in the Bayesian method.

Table 1. Multicollinearity between parameters

	ET	IT	MP	LW	VIF
ET	1	0,935	0,197	0,710	138,455
IT		1	0,157	0,694	102,578
MP			1	0,142	36,148
LW				1	195,748

Since the bend point parameter "e" is examined, it takes values between -0,112 and -0,119, in the LS method, and between -0,119 and -0,117 in the Bayesian method. While the lowest bending point parameter value is observed in Simmental with the values of -0,119 and -0,117 in the highest value is in both methods.

Additionally; the explanatory variance (R^2) values, which shows the highest value takes between 0,929 and 0,921 in its method, while in the Bayesian method it has values between 0,929 and 0,919. This value was followed for Simmental by 0,908 in LS method and 0,919 in Bayesian method. The highest R^2 value was observed in Holstein with the values of 0,921 in the LS method and 0,929 in the Bayesian method.

Table 2. Estimated parameter values according to Least Squares and Bayesian method and sum of error squares

Parameter	Estimate values					
	Simmental		Holstein		EAR	
	LS	Bayesian	LS	Bayesian	LS	Bayesian
a	5313,738	5311,575	5364,142	5360,244	5326,648	5313,986
b (ET)	1,758	1,751	1,658	1,647	1,708	1,698
c (IT)	37,470	37,578	36,363	36,912	36,812	37,014
d (MP)	123,134	122,143	125,452	123,550	124,286	124,584
e (LW)	-0,119	-0,117	-0,112	-0,117	-0,118	-0,115
R^2	0,908	0,919	0,921	0,929	0,916	0,918
Sum of Error Squares						
LS	1672,756	1789,812	1842,816	1835,352	1714,785	1666,815
Bayesian	1692,874	1795,312	1866,620	1846,544	1726,387	1669,182

Acknowledgement

This study was presented as oral presentation at the International Congress on Domestic Animal Breeding Genetics and Husbandry (ICABGEH 2020), 12-14 August 2020, Samsun/ TURKEY.

Conflict of interest

The author declares that there is no conflict of interest.

References

Atkinson, A. C. (2000). *Robust diagnostic regression analysis*. McGraw-Hill Comp.

Breusch, T. S. (1978). Testing for Autocorrelation in Dynamic Linear Models. *Australian Economic Papers*, *17*(31), 334-355. https://doi.org/10.1111/j.1467-8454.1978.tb00635.x

Draper, N. R., & Smith, H. (1981). *Applied regression analysis* (2nd ed.). John Wiley & Sons.

Durbin, J. (1970). Testing for Serial Correlation in Least Squares Regression when some of the Regressors are Lagged Dependent Variables. *Econometrica*, *38*(3), 410-421. https://doi.org/10.2307/1909547

Erdem, H., Atasever, S., & Kul, E. (2007). Gökhöyük Tarim İşletmesinde Yetiştirilen Siyah Alaca Sığırların Süt ve Döl Verim Özellikleri 2. Döl Verim Özellikleri. *Ondokuzmayıs Üniversitesi Ziraat Fakültesi Dergisi*, *22*(1), 47-54. (in Turkish)

Ferreira, W. J., Texeira, N. M., Torres, R. de A., & da Silva, M. B. (2002). Utilização da Produção de Leite No Dia do Controle na Avaliação Genéticaem Gado de Leite - Uma Revisão. *Archivos Latinoamericanos de Producción Animal*, *10*(1), 46-53. (in Spanish)

Jensen, J. (2001). Genetic Evaluation of Dairy Cattle Using Test-Day Models. *Journal of Dairy Science*, *84*(12), 2803-2812. https://doi.org/10.3168/jds.S0022-0302(01)74736-4

Leamer, E. E. (1973). Multicollinearity: A Bayesian Interpretation. *The Rewiev of Economics and Statistics*, *55*(3), 371-380. https://doi.org/10.2307/1927962

Ljung, G. M., & Box, G. E. P. (1978). On a Measure of Lack of Fit in Time Series Models. *Biometrika*, *65*(2), 265-270. https://doi.org/10.1093/biomet/65.2.297

Montgomery, D. C., & Peck, E. A. (1992). *Introduction to linear regression analysis*. John Wiley & Sons.

Rao, C. R. (1973). *Linear statistical inference and its application* (2nd ed.). Wiley.

Rekaya, R., Carbaño, M. J., & Toro, M. A. (1999). Use of Test Day Yield for The Genetic Evaluation of Production Traits in Holstein-Friesian Cattle. *Livestock Production Science*, *57*(3), 203-217. https://doi.org/10.1016/S0301-6226(98)00181-X

Vila, J. P., Wagner, V., & Neveu, P. (2000). Bayesian Nonlinear Model Selection and Neural Networks: A Conjugate Prior Approach. *IEEE Transactions on Neural Networks*, *11*(2), 265-78. https://doi:10.1109/72.838999

Visscher, P. M., & Goddard, M. E. (1995). Genetic Parameters for Milk Yield, Survival, Workability, and Type Traits for Australian Dairy Cattle. *Journal of Dairy Science*, *78*(1), 205-220. https://doi.org/10.3168/jds.S0022-0302(95)76630-9

Zellner, A. (1971). *An introduction to Bayesian inference in econometrics*. John Wiley.