RESEARCH ARTICLE
Journal of Biometry Studies (2024) 4(1): 31-41
DOI: 10.61326/jofbs.v4i1.03
www.biometrystudies.com

Journal of Biometry Studies

PRENSIP
www.prensip.gen.tr

Evaluation of the effect of morphological traits on fish growth by comparison using ridge and ordinary least squares regression

Mehmet Fatih CAN[1,*], Cemil KARA[2]

[1]İskenderun Technical University, Faculty of Marine Science and Technology, Hatay/TÜRKİYE
[2]Karadeniz Technical University, Faculty of Science, Department of Biology, Trabzon/TÜRKİYE

[*]Corresponding author: mfatih.can@iste.edu.tr Received: 26/04/2024, Accepted: 21/06/2024

Abstract

Growth in fish is characterized by length and weight and studies that encompassing relationships among the fish length and weigh with some morphometric traits provide crucial information in the field of fish biology. High correlations (or multicollinearity) among the morphometric traits in fish morphology studies is a well-known phenomenon. If the relationship between growth and morphometry is modeled using the ordinary least squares estimator (OLS), the parameter estimates are likely to be too large in absolute value and possibly have the wrong sign due to the problem of multicollinearity. The ridge regression (RR) estimator has been proposed to avoid the adverse effects of multicollinearity among the regressors. In this study, therefore, multiple linear RR was used to model relationship fish length and fish weight to some predictive metric traits. Predictive traits were predorsal length, head length, post dorsal length, head height, and eye diameter. Data were derived from a total of 126 *Capoeta damascina* individuals sampled from the Euphrates, Türkiye. The ridge optimal k-parameter, producing acceptable variance influence factor (VIF<10), was determined as 0.021 for both of growth indices considering a combination of ridge trace and VIF trace plots and MSE and VIF values produced from an array of k values ($0.0 \leq k \leq 1.00$). Although, small decreases were observed in the adjusted *R*-square values (ARS) obtained by RR compared to the ARS values obtained by OLS, a significant decrease in VIF values outweighs this drawback, indicating that the models are more stable. The methodological approach and findings in this study may contribute to filling a gap in the literature regarding the relationships between fish growth and morphology. Additionally, it could enhance better growth predictions for different fish species, aiding sustainable fisheries management and the selective cultivation of desired fish traits in aquaculture by improving the understanding of morphometric features.

Keywords: Fish morphometry, Multicollinearity, Variance influence factor, *Capoeta damascina*

Please cite this article as follows:

Can, M. F., & Kara, C. (2024). Evaluation of the Effect of Morphological Traits on Fish Growth by Comparison Using Ridge and Ordinary Least Squares Regression. *Journal of Biometry Studies*, *4*(1), 31-41. https://doi.org/10.61326/jofbs.v4i1.03

1. Introduction

Fish morphometric traits are quantifiable characteristics related to the shape, size, and proportions of fish bodies. Morphometric studies encompassing these relationships provide crucial information in the field of fish biology, including species growth, feeding behavior, ecological strategies, niche partitioning, habitat utilization, and trophic structure. For instance, examining the linear relationships between total length, body height, and width reveals growth patterns, the height of the caudal peduncle influences swimming efficiency and overall fitness, while pectoral fin length and eye diameter reflect adaptations related to habitat, feeding, and locomotion. These morphometric analyses contribute significantly to our understanding of fish diversity and ecological dynamics (Clabaut et al., 2007; Mojekwu & Anumudu, 2015; Keppeler et al., 2020; Pino et al., 2021). In addition to these, morphometric features can also be used in aquaculture to distinguish between fish escaped from the rearing environment and wild fish (Arechavala-Lopez et al., 2012; Lenhardt et al., 2012).

Morphometric characteristics are taken into account in stock discrimination studies in fisheries science using multivariate statistical methods such as Principal Component Analysis, Cluster Analysis, and Linear Discriminant Analysis (Winans, 1987; Turan, 1999; Mojekwu & Anumudu, 2015). But, due to allometric growth that is an inherent characteristic of fishes which is a common problem with morphometric data is that all measurements are highly correlated with length, the morphometric traits are required to be transformed before analysis in stock discriminant studies. This high correlation between morphometric features is called multicollinearity. Due to this situation, the simple linear regression method is generally used to evaluate the relationships between fish growth and morphometric characteristics, except for stock discrimination studies (Kyritsi & Moutopoulos, 2018; Orbach et al., 2019; Yazıcı & Yazıcıoğlu, 2020).

In cases where there is multicollinearity; (i) The values of parameters estimated by the Ordinary Least Squares (OLS) method may differ significantly from their actual values, (ii) Regression coefficients are uncertain and the standard errors of these coefficients are infinite, (iii) Variance and covariances of the regression coefficients are tend to increase, (iv) *R*-square value of the model is high, However, none or very few of the independent variables may be significant according to the partial *t*-test, and (v) The direction of the relationships of the relevant independent variables with the dependent variable may contradict theoretical and experimental expectations. Although there are statistical approaches such as principal component analysis (PCA) and removing the variable having VIF higher than 10 are used to overcome this problem, the use of these approaches may lead to information loss. Another approach to overcome this problem is to use ridge regression. Ridge regression is a statistical method used to estimate coefficients in multiple regression models when independent variables are highly correlated without removing any variables from model (Albayrak, 2005; Karakaş, 2008; McDonald, 2009; Büyükuysal & Öz, 2016; Nandi & Saikia, 2015; Özkale & Altuner, 2021).

Although ridge regression has been used for different purposes in fish related studies (Rikardsen & Johansen, 2003; Okamura et al., 2017; Niu et al.,2019; Martins et al., 2021; Liyandja et al., 2022; Gao et al., 2024), it seems that it has not been used to describe the relationship between fish morphology and fish growth to date. In essence, the fish material used in this study were considered for example purposes only to illustrate why and how ridge regression is used in modeling the relationship between fish growth and morphometry. Therefore, the main purpose of this study is to fill a gap in the literature on this subject by holistically estimating the relationship between total fish length and total weight characteristics and a group of morphometric features using the ridge regression method. Moreover, the methodological approach in the study can contribute to (i) sustainable fisheries management by making better growth predictions for different fish species and (ii) a better understanding.

2. Material and Methods

2.1. Fish material

The morphological features used in this study was belong to the *Capoeta damascina* species. The specimen measured in this study were provided from a research project that was supported by the Adıyaman University Scientific Research Projects (BAP) Coordination Unit, Project No. 2012/001, and was carried out with the legal consent of the Ministry of Food, Agriculture, and Livestock, General Directorate of Fisheries and Aquaculture (permission date 05.04.2012; permission number: 01515). The *C. damascina* species group occurs in the entire Levant, Mesopotamia, the Orontes, Iran, and the southern and eastern parts of Türkiye (Alp et al., 2005; Alp et al., 2013; Asadollah et al., 2017). A total of 126 *C. damascina* specimens caught from the Adıyaman region (in the middle Euphrates Basin) of Türkiye between April 2012 and December 2013 were caught in streams by using electroshock devices. Then, all length related measurements (mm) were made with a digital caliper with a precision of 0.01 mm, and the weight measurements were made with a digital scale with a precision of 0.01 g. Since this study was based on a modeling methodology approach, there was no need to provide further information about the fish and the area where they were sampled.

2.2. Metric traits

In the study total lenght (TL), total weight (TW), body height (BH), predorsal length (PR), head length (HL), post dorsal length (PD), head height (HH) and eye diameter (ED) metric features were taken into consideration.

2.3. Statistical framework

2.3.1. The general framework of the study

In the study total lenght (TL), total weight (TW), body height (BH), predorsal length (PR), head length (HL), post dorsal length (PD), head height (HH) and eye diameter (ED) metric features were taken into consideration.

Irrespective of used domain, the biggest debate with the ridge estimator is the selection of the regularization parameter (*k*). Therefore, the proposed general framework of the study was given below in hierarchical format:

a. Calculation of some descriptive statistics of all variables

b. Calculation and visualization of the correlation matrix between variables: It will give us a preliminary idea about whether there is a multicollinearity problem.

c. Calculation of VIF values for predictor variables based on Ordinary Least Square (OLS): If there is a doubt determined in step “b”, VIF values, which are indicative of multicollinearity, are calculated for each predictor variable. The presence of variables with VIF equal and greater than 10 quantitatively indicates the presence of multicollinearity.

d. To have an idea about the approximate optimum k-parameter visually using graphical methods (VIF Trace and Ridge Trace Plot) for a range of the Ridge k parameter.

e. Calculate the models' minimum Mean Square Error (MSE), minimum generalized cross-validation (GCV) and adjusted *R*-squared values for a range of Ridge k parameters.

f. Calculate the VIF values of the predictor variables for a range of the ridge k-parameter.

g. Deciding on the optimum ridge k value by considering the smallest MSE, adj-*R*-square and VIF values in a balanced manner.

2.3.2. Models

In this study TL and TW were considered dependent and other metric traits were considered independent (regressor) variables. Regardless relationship between dependent and independent variables, to simplify the modelling dependent variables were modelled with multiple linear regression approach as following:

$$TL = \alpha + \beta_1 PR + \beta_2 PD + \beta_3 HH + \beta_4 BH + \beta_5 HL + \beta_6 ED + \varepsilon$$

$$TW = \alpha + \beta_1 PR + \beta_2 PD + \beta_3 HH + \beta_4 BH + \beta_5 HL + \beta_6 ED + \varepsilon$$

where, α, β, and ε are intercept, regression coefficients, and error term, respectively.

2.3.3. Multicollinearity and variance inflation factor (VIF)

In regression analysis, some assumptions are taken into account when estimating the parameters that show the relationship between dependent and independent variables. In order to determine the relationship between variables with multiple regression analysis, the assumptions of the multiple linear regression model must be met. However, the assumptions of the multiple regression model are not valid in the analysis of every event. The assumptions in question are the assumptions about the error vector and the independent variable matrix. One of these assumptions that ensures that the standard errors of the estimates are smallest and therefore the parameter estimates are effective is that there is no relationship between the independent variables. This is called multicollinearity. As can be seen from the graphs above (Figure 1 and Figure 2), the existence of high correlations between the predictor variables is a sign of this situation. If there is a perfect correlation between the explanatory variables, that is, if the correlation coefficient for these variables is equal to 1, the parameters become undeterminable. It becomes impossible to find separate numerical values for each parameter, and in this case the least squares method cannot be used. On the other hand, if there is no correlation between the explanatory variables, that is, if the correlation coefficient for these variables is equal to 0, they are called orthogonal variables and do not pose a problem in estimating the coefficients. Multicollinearity may occur for various reasons (Katz, 2006; Adeboye et al., 2014; Daoud, 2017). However, considering the current study, the main reason for multicollinearity is that the variables of interest tend to change together over time.

One of the approaches used to detect multicollinearity is the variance inflation factor (VIF). The variance inflation factor is calculated to determine the degree of relationship of an independent variable with other independent variables. If there is no relationship between the dependent variable and the independent variables (*R*-square=0), the variance inflation factor is equal to 1 (1/(1- *R*-square)=1/(1-0)=1). If there is a perfect relationship between the dependent and independent variables (*R*-square=1), the variance inflation factor [VIF=1/(1- *R*-square)=1/(1-1)] will be infinite. If *R*-square=90%, the variance inflation factor is obtained as 10 [VIF=1/(1-0.90)=10]. In general, the following general rule applies for interpretation: if the variance inflation factor is equal to or greater than 10 (VIF>10), there is a multicollinearity problem (Thompson et al., 2017; Oke et al., 2019). In case of collinearity, it is necessary to eliminate it in order to reduce the standard error of coefficient estimates that determine the relationship between dependent and independent variables and to make more consistent estimates. One of the methods used in case of multicollinearity is Ridge regression.

2.3.4. Ridge regression

The solution technique of ridge regression (RR) is similar to the ordinary least squares (OLS) method. The RR method, is performed by adding a small and positive constant to the diagonal elements of the ($X^T X$) matrix formed by the variables in standard form before calculating the regression coefficient estimates. Accordingly, the ridge regression solution is as follows;

$$\hat{\beta}_{RR} = (X^T X + kI)^{-1} X^T Y$$

In the formula "I" and "k" represent the identity matrix and the ridge parameter, respectively. The difference of the RR method from OLS is the presence of the k-ridge parameter. Since the ridge solution for k=0 is equivalent to the OLS solution, the ridge estimate can also be expressed as a linear transformation of the least squares estimate. In the RR method, the steps followed in the least squares method are repeated more than once. Among the parameter estimates calculated for each k that has a value between 0 and 1, those that meet the sought criterion are determined.

One of the main obstacles in using RR is in choosing an appropriate value of k. The inventors of RR suggested using a graphic which they called the ridge trace. This plot shows the RR coefficients as a function of *k*. When viewing the ridge trace, the analyst picks a value for k for which the regression coefficients have stabilized. Often, the regression coefficients will vary widely for small values of k and then stabilize. Choose the smallest value of k possible (which introduces the smallest bias) after which the regression coefficients seem to remain constant. Note that increasing k will eventually drive the regression coefficients to zero.

2.4. Analysis tools

In the study, the packages "lmridge" (Ullah et al., 2018), "psych" (Revelle & Revelle, 2015), and "rattle" (Williams, 2011) in the R program were used.

lmridge contains functions related to fitting of the RR model and provides a simple way of obtaining the estimates of RR coefficients, testing of the ridge coefficients, and computation of different ridge related statistics, which prove helpful for selection of optimal biasing parameter *k*. The four arguments of lmridge() function are described in Table 2 (Ullah et al., 2018). The syntax of default function is,

lmridge (formula, data, scaling = ("sc","scaled","centered"),k,...)

Table 1. Description of lmridge() function arguments

Argument	Description
Formula	Symbolic representation for RR model of the form, response ~ predictors.
Data	Contains the variables that have to be used in RR model.
k	The biasing parameter, may be a scalar or vector. If a *k* value is not provided, *k*=0 will be used as the default value, i.e., the OLS results will be produced.
Scalling	The methods for scaling the predictors. The sc option uses the default scaling of the predictors in correlation form; the *scaled* option standardizes the predictors having zero mean and unit variance; and the *centered* option centers the predictors.

There are many scientific approaches to choosing the correct ridge parameter. The ultimate choice of *k* for a particular application involving linear explanatory variables still remains part art, part science (Forrester & Kalivas, 2004; Khalaf & Shukur, 2005; McDonald, 2009; Mami et al., 2021). In this study, in deciding on the *k* parameter, a series of *k* parameter (0.0≤*k*≤1.0) was considered into iterative manner in each of steps described under general framework of the study.

3. Results

3.1. Descriptive statistics and correlations

Some statistics (mean, standard deviation, coefficient of variation) of the metric features taken into account in the research are given in Table 1. Cross-correlations between variables, histogram distributions of the variables, and Correlation dendrogram are given in Figure 1 and Figure 2, respectively. In terms of variability (measured by Cv, %), TW had the highest (84.09%) as a nature of weight, others somehow were similar except ED (14.97%).

Table 2. Mean, standard deviation (Sd) and coefficient of variation (CV, %) values of the variables

Metric trait	Mean	Sd	CV, %
ED	6.13	0.92	14.97
BH	34.96	8.82	25.22
HH	21.70	6.10	28.12
HL	32.48	8.07	24.86
PD	54.38	16.15	29.71
PR	68.61	17.04	24.84
TL	171.16	45.04	26.32
TW	58.58	49.26	84.09

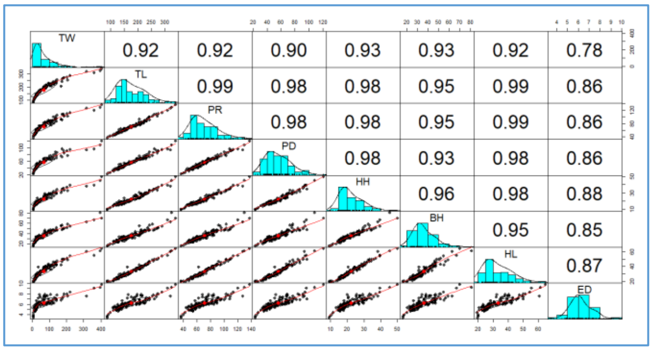

Figure 1. Cross correlations and linear relationships between variables, and histogram distribution of each variable

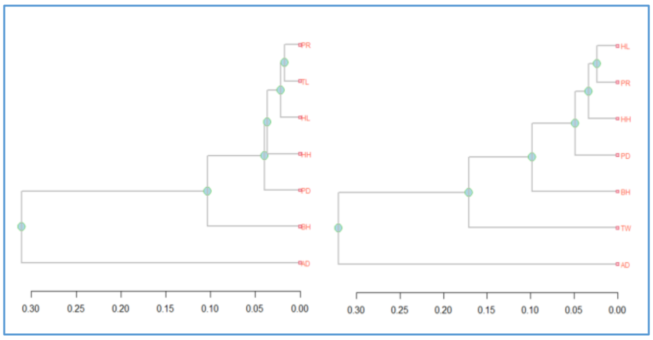

Figure 2. Correlation dendrogram graphs of variables (Total height (TL) on the left-hand side and total weight (TW) on the right-hand side)

As can be seen in Figure 1 and Figure 2, there are quite high correlations between the predictor variables (ED, BH, HH, HL, PD, and PR), and between predictors and dependent variables (TL and TW). High correlations observed between independent variables indicate a potential multicollinearity problem. However, since simple correlation coefficients alone are not sufficient, it is not possible to be sure of the existence of a multicollinearity problem (Ölmez & Yağanoğlu, 2023).

3.2. Ordinary Least Square (OLS) results and VIF values

Ridge solution for *k*=0 is equivalent to the ordinary least squares solution (OLS). In case *k*=0 for TL, the RR analysis resulted that adj-*R*-squared=0.984, MSE=8626.853, and regression coefficients of PR, PD, BH, and HL were significant (*p*<0.05), whereas HH and ED were not (*p*>0.05). For TW, adj-*R*-squared=0.886, MSE=127826.47, and only regression coefficients of BH and ED were significant (*p*<0.05), in contrast other predictors were not (*p*>0.05) (Figure 3, Figure 4).

```
Coefficients: for Ridge parameter K= 0
           Estimate Estimate (Sc) StdErr (Sc) t-value (Sc) Pr(>|t|)
Intercept   -4.0265   -31703.5376   4563.6421      -6.9470   <2e-16 ***
PR           0.9487      212.8959     49.6621       4.2869   <2e-16 ***
PD           0.7481      155.0240     38.3154       4.0460   0.0001 ***
HH          -0.9275      -75.0313     43.1890      -1.7373   0.0848 .
BH           0.4487       54.4057     24.6356       2.2084   0.0291 *
HL           2.2365      237.3432     45.1410       5.2578   <2e-16 ***
ED           0.1902        2.2332     13.5112       0.1653   0.8690
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

     R2    adj-R2  DF ridge          F       AIC        BIC
0.98540   0.98480   5.99995 1391.39510 487.13857 1137.12311

           PR       PD       HH       BH       HL      ED
 k=0 60.84427 36.21742 46.01677 14.97258 50.27037 4.50356
```

Figure 3. Regression summary and VIF values of RR for *k*=0 that equals to results of ordinary least squares for TL relation to predictors

```
Coefficients: for Ridge parameter K= 0
           Estimate Estimate (Sc) StdErr (Sc) t-value (Sc) Pr(>|t|)
Intercept -144.7147   -29911.3869  17566.9195      -1.7027   0.0911 .
PR           0.9802      219.9666    191.1652       1.1507   0.2521
PD          -1.0923     -226.3406    147.4884      -1.5346   0.1274
HH           3.0434      246.1986    166.2483       1.4809   0.1412
BH           3.0943      375.2203     94.8304       3.9568   0.0001 ***
HL           2.4827      263.4704    173.7621       1.5163   0.1320
ED          -9.2464     -108.5916     52.0088      -2.0879   0.0389 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Ridge Summary
       R2    adj-R2  DF ridge         F       AIC        BIC
  0.89080   0.88640   5.99995 168.53505 837.59178 1487.57632

             PR       PD       HH       BH       HL      ED
   k=0 60.84427 36.21742 46.01677 14.97258 50.27037 4.50356
```

Figure 4. Regression summary and VIF values of RR for *k*=0 that equals to results of ordinary least squares for TW relation to predictors

VIF values of predictor variables of both TL and TW based on ordinary least squares (OLS) were calculated as; PR=60.84, PD=36.22, HH=46.02, BH=14.97, HL=50.27, and ED=4.50. As suspected above from correlation results, VIF values clearly show that there was multicollinearity among the predictor variables for both of TL and TW (Figure 3 and Figure 4).

3.3. Searching for optimal "*k*" value

3.3.1. Graphical Searching

In graphical searching, the ridge trace plots and VIF trace plot for a range of k parameter (0.0≤*k*≤1.0) for TL and TW estimations were given in Figure 5, respectively. Ridge Trace Plot shows the effect of different *k*-ridge values on the regression coefficient of each predictor variable for dependent variable (TL and TW). VIF Trace plot shows the effect of different *k*-ridge values on the VIF values of each predictor variable for the dependent variable (TL and TW). As can be seen from the "Ridge Trace Plot" graphs for TL and TW, as the *k* values increase, the "Ridge" beta coefficients get closer to each other. On the graph, it is reported that the smallest MSE (7131.745) value for TL was reached at *k*=0.003. For TW, the lowest MSE value (59893.122) was reached at *k*=0.016.

VIF Trace plot shows the decrease in VIF values of predictor variables despite increasing *k* values according to generalized cross-validation or GCV criterion (Golub et al., 1979). The minimum GCV value for TL was reached at *k*=0.003, and the minimum GCV value for TW was reached at *k*=0.018.

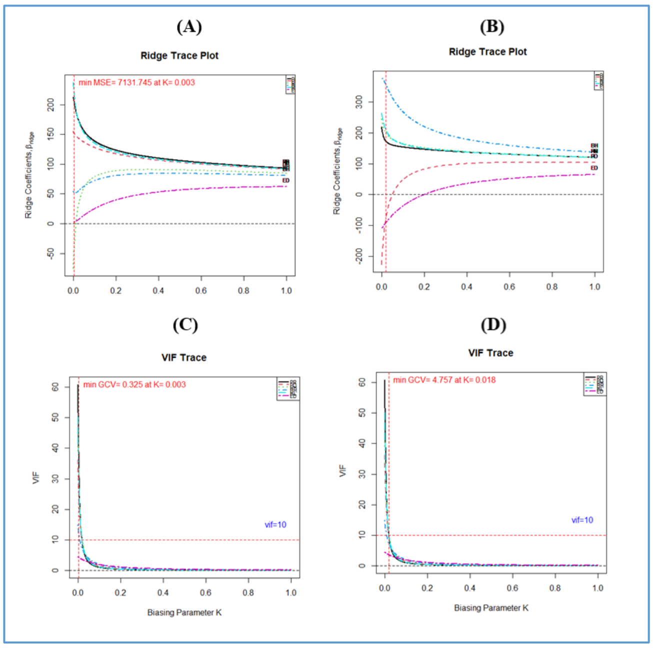

Figure 5. Ridge trace plots {TL-(A), TW-(B)} and VIF trace plot {TL-(C), TW-(D)} for a range of *k* parameter (0.0≤*k*≤1.0) for TL and TW estimations

3.3.2. Searching of the best *k* by minimum MSE

According to the minimum MSE approach, the ridge *k* value that produces the least MSE value (7131.745) for TL prediction is 0.003 and the adj-*R*-square value of the model is 0.9825 (Figure 6), while for TW prediction these values are found to be 59893.12, 0.016 and 0.8722, respectively (Figure 7).

```
Ridge Regression Statistics 1:

          Variance     Bias^2       MSE         F     R2 adj-R2  rsigma2
K=0      8626.8525     0.0000  8626.853 1391.3951 0.9854 0.9848  40.5350
K=0.001 7597.0579   141.9010  7738.959 1390.7822 0.9846 0.9840  40.5528
K=0.002 6770.2743   510.1914  7280.466 1389.1517 0.9839 0.9832  40.6004
K=0.003 6093.2526  1038.4927  7131.745 1386.7571 0.9832 0.9825  40.6705
K=0.004 5529.6500  1679.5794  7209.229 1383.7914 0.9826 0.9819  40.7577
K=0.005 5053.8526  2399.2746  7453.127 1380.4063 0.9819 0.9812  40.8576
K=0.006 4647.3440  3172.5233  7819.867 1376.7147 0.9813 0.9806  40.9672
K=0.007 4296.4114  3980.7806  8277.192 1372.8044 0.9808 0.9800  41.0839
K=0.008 3990.6959  4810.2259  8800.922 1368.7428 0.9802 0.9794  41.2058
K=0.009 3722.2367  5650.5154  9372.752 1364.5802 0.9796 0.9788  41.3315
K=0.01  3484.8110  6493.8924  9978.703 1360.3576 0.9791 0.9783  41.4598
K=0.011 3273.4960  7334.5447 10608.041 1356.1045 0.9786 0.9777  41.5898
K=0.012 3084.3390  8168.1326 11252.472 1351.8459 0.9780 0.9772  41.7208
K=0.013 2914.1386  8991.4389 11905.577 1347.5998 0.9775 0.9766  41.8523
K=0.014 2760.2772  9802.1063 12562.383 1343.3779 0.9770 0.9761  41.9838
K=0.015 2620.5856 10598.4394 13219.025 1339.1917 0.9765 0.9756  42.1151
K=0.016 2493.2558 11379.2533 13872.509 1335.0498 0.9761 0.9751  42.2457
K=0.017 2376.7760 12143.7575 14520.533 1330.9557 0.9756 0.9746  42.3757
K=0.018 2269.8610 12891.4668 15161.328 1326.9148 0.9751 0.9741  42.5047
K=0.019 2171.4208 13622.1315 15793.552 1322.9292 0.9746 0.9736  42.6328
K=0.02  2080.5180 14335.6835 16416.201 1319.0028 0.9742 0.9731  42.7597
K=0.021 1996.3521 15032.1940 17028.546 1315.1340 0.9737 0.9726  42.8855
K=0.022 1918.2283 15711.8402 17630.069 1311.3225 0.9733 0.9722  43.0101
K=0.023 1845.5362 16374.8796 18220.416 1307.5719 0.9728 0.9717  43.1335
```

Figure 6. MSE, adj-*R*-square and other statistics produced by different *k* values for TL estimation

```
Ridge Regression Statistics 1:

            Variance     Bias^2       MSE        F     R2 adj-R2  rsigma2
K=0      127826.4729     0.0000 127826.47 168.5350 0.8908 0.8864 600.6178
K=0.001  112516.8563   282.9445 112799.80 168.5370 0.8897 0.8852 600.6110
K=0.002  100153.0506  1019.8292 101172.88 168.5388 0.8886 0.8841 600.6044
K=0.003   89983.5825  2082.7516  92066.33 168.5367 0.8876 0.8831 600.6119
K=0.004   81489.7399  3381.6844  84871.42 168.5286 0.8867 0.8821 600.6409
K=0.005   74302.2218  4851.7265  79153.95 168.5140 0.8858 0.8812 600.6930
K=0.006   68151.6381  6445.2066  74596.84 168.4925 0.8849 0.8802 600.7694
K=0.007   62836.8661  8126.5912  70963.46 168.4644 0.8840 0.8793 600.8696
K=0.008   58204.8844  9869.0876  68073.97 168.4299 0.8832 0.8785 600.9927
K=0.009   54137.3490 11652.3121  65789.66 168.3892 0.8824 0.8776 601.1380
K=0.01    50541.2370 13460.6529  64001.89 168.3428 0.8816 0.8768 601.3036
K=0.011   47342.5983 15282.0960  62624.69 168.2911 0.8808 0.8760 601.4886
K=0.012   44481.8079 17107.3689  61589.18 168.2345 0.8801 0.8752 601.6908
K=0.013   41910.3969 18929.3079  60839.70 168.1735 0.8793 0.8744 601.9091
K=0.014   39588.6056 20742.3832  60330.99 168.1082 0.8786 0.8737 602.1429
K=0.015   37483.4033 22542.3398  60025.74 168.0391 0.8779 0.8730 602.3906
K=0.016   35567.2001 24325.9223  59893.12 167.9666 0.8772 0.8722 602.6507
K=0.017   33816.8752 26090.6628  59907.54 167.8907 0.8765 0.8715 602.9231
K=0.018   32212.7526 27834.7164  60047.47 167.8119 0.8758 0.8708 603.2062
K=0.019   30738.1179 29556.7318  60294.85 167.7303 0.8752 0.8701 603.4995
K=0.02    29378.5915 31255.7502  60634.34 167.6465 0.8745 0.8695 603.8013
K=0.021   28121.8764 32931.1248  61053.00 167.5603 0.8739 0.8688 604.1119
K=0.022   26957.2962 34582.4571  61539.75 167.4718 0.8732 0.8681 604.4311
K=0.023   25875.4768 36209.5461  62085.02 167.3817 0.8726 0.8675 604.7564
```

Figure 7**.** MSE, adj-*R*-square and other statistics produced by different *k* values for TW estimation

3.3.3. Searching of the best *k* by acceptable VIF

Figure 8 gives the VIF values of each estimator variable for different *k*-ridge parameter values for TL and TW modelling. In contrast to the k values increased, VIF values were values decreased. The first *k* value where the VIF values of the predictor variables were all calculated below 10 was found to be 0.021.

3.3.4. Final decision on the optimum *k* parameter

So far, to decide the optimum k parameter, we have used the graphical method, calculating MSE, VIF and *R*-squared values. In graphical methods, we considered Ridge Trace Plot and VIF Trace Plot. In the Ridge Trace Plot, the *k* value that produces the lowest MSE is also given on the graph. For TL, *k*=0.003 and MSE=7131.745, for TW, *k*=0.016 and MSE=59893.122 values were produced. In VIF Trace Plot, the *k* value that produces the minimum GCV value is also given on the graph. While the *k* value that produced the lowest GCV for TL (0.325) was 0.003, the *k* value that produced the minimum GCV for TW (4.757) was found to be 0.018.

```
> vif(mod)
              PR       PD       HH       BH       HL      ED
k=0     60.84427 36.21742 46.01677 14.97258 50.27037 4.50356
k=0.001 51.68972 32.35179 41.03699 14.06270 43.77091 4.42522
k=0.002 44.53242 29.15939 36.86107 13.28556 38.56067 4.35466
k=0.003 38.82066 26.47946 33.31653 12.61041 34.30272 4.29006
k=0.004 34.18275 24.19909 30.27697 12.01567 30.76668 4.23016
k=0.005 30.36048 22.23654 27.64743 11.48566 27.79003 4.17403
k=0.006 27.16968 20.53107 25.35509 11.00873 25.25501 4.12101
k=0.007 24.47595 19.03654 23.34313 10.57601 23.07430 4.07060
k=0.008 22.17925 17.71724 21.56655 10.18061 21.18183 4.02241
k=0.009 20.20378 16.54507 19.98926  9.81710 19.52677 3.97614
k=0.01  18.49121 15.49760 18.58198  9.48111 18.06934 3.93154
k=0.011 16.99604 14.55672 17.32073  9.16911 16.77804 3.88843
k=0.012 15.68230 13.70761 16.18569  8.87818 15.62758 3.84665
k=0.013 14.52127 12.93807 15.16034  8.60590 14.59747 3.80607
k=0.014 13.48974 12.23792 14.23081  8.35026 13.67089 3.76658
k=0.015 12.56881 11.59862 13.38540  8.10952 12.83397 3.72809
k=0.016 11.74294 11.01296 12.61416  7.88222 12.07513 3.69053
k=0.017 10.99925 10.47480 11.90857  7.66711 11.38466 3.65382
k=0.018 10.32701  9.97889 11.26132  7.46308 10.75433 3.61792
k=0.019  9.71720  9.52069 10.66610  7.26919 10.17718 3.58276
k=0.02   9.16218  9.09630 10.11745  7.08461  9.64721 3.54832
k=0.021  8.65547  8.70230  9.61059  6.90859  9.15929 3.51454
k=0.022  8.19152  8.33573  9.14135  6.74049  8.70896 3.48140
k=0.023  7.76558  7.99398  8.70608  6.57973  8.29238 3.44886
k=0.024  7.37353  7.67474  8.30156  6.42578  7.90617 3.41690
k=0.025  7.01182  7.37600  7.92495  6.27819  7.54738 3.38550
k=0.026  6.67733  7.09595  7.57371  6.13654  7.21341 3.35462
k=0.027  6.36735  6.83299  7.24561  6.00043  6.90199 3.32426
k=0.028  6.07951  6.58570  6.93865  5.86954  6.61108 3.29438
```

Figure 8. Effect of different *k*-ridge values on the VIF value of each estimator variable for the total length (TL) and total weight (TW). Since VIF values decreased significantly, no further evaluation on *k* values were presented in the figure

When we searched for the minimum MSE values for different *k* values one by one, it was seen that the *k* value that produced the minimum MSE value was the same as the Rigge Trace Plot for TL and TW predictions. However, the first *k* value where each predictor variable in the model met the condition of a VIF value less than 10 was found to be *k*=0.021 for both predictions. In this case, it will be necessary to choose between the values of *k*=0.003 and *k*=0.0214 for TL and the values of *k*=0.016, *k*=0.018 and *k*=0.0214 for TW.

Although the lowest MSE was obtained when *k*=0.003 was selected for TL (7131.745), the VIF value of no predictor variable, except for the ED predictor variable (4.29), did not fall below the acceptable threshold value (VIF<10). On the other hand, if *k*=0.021 is selected, the MSE will increase slightly (17028.546) and the adj-*R*-square value will decrease slightly (from 0.9825 to 0.9726), and the VIF values of the predictor variables will be reduced to acceptable limits (VIF<10) (Figure 9).

```
> vif(mod)
              PR       PD       HH       BH       HL      ED
k=0.003 38.82066 26.47946 33.31653 12.61041 34.30272 4.29006
k=0.021  8.65547  8.70230  9.61059  6.90859  9.15929 3.51454
> rstats1(mod)

Ridge Regression Statistics 1:

        Variance    Bias^2       MSE        F     R2 adj-R2 rsigma2       CN
K=0.003 6093.253  1038.493  7131.745 1386.757 0.9832 0.9825 40.6705 420.4031
K=0.021 1996.352 15032.194 17028.546 1315.134 0.9737 0.9726 42.8855 180.6727
```

Figure 9. VIF (VIF) values and some statistics (rstats1) for *k*={0.003, 0.021} values for TL estimation

If k=0.016 and k=0.018 are selected for TW, lower MSE (59893.12 and 60047.47) and higher adj-R-square (0.8722 and 0.8708) values can be obtained. However, when k=0.016, the VIF values of the PR, PD, HH and HL predictor variables are greater than the acceptable threshold value (VIF>10). On the other hand, if k=0.021 is selected, although the MSE values are slightly high (61053.00) and the adj-R-square (0.8688) value is slightly low, the VIF values of the predictor variables drop to an acceptable level (VIF<10) (Figure 10).

```
> vif(mod)
              PR       PD       HH       BH       HL       ED
k=0.016 11.74294 11.01296 12.61416 7.88222 12.07513 3.69053
k=0.018 10.32701  9.97889 11.26132 7.46308 10.75433 3.61792
k=0.021  8.65547  8.70230  9.61059 6.90859  9.15929 3.51454
> rstatsl(mod)

Ridge Regression Statistics 1:

        Variance   Bias^2      MSE        F     R2 adj-R2  rsigma2       CN
K=0.016 35567.20 24325.92 59893.12 167.9666 0.8772 0.8722 602.6507 214.5852
K=0.018 32212.75 27834.72 60047.47 167.8119 0.8758 0.8708 603.2062 199.5919
K=0.021 28121.88 32931.12 61053.00 167.5603 0.8739 0.8688 604.1119 180.6727
```

Figure 10. VIF (VIF) values and some statistics (rstatsl) for k={0.016, 0.018, 0.021} values for TW estimation

Considering all the above findings, larger k values have the potential to both bring the regression coefficients closer to zero and produce lower adj-R-squared values. In addition, lower k values have the potential to produce higher VIF values and some potentially important predictor variables to be insignificant. Therefore, it was decided to take the optimum ridge k value as 0.021 for TL and TW predictions.

3.4. Ridge regression analysis results

The ridge regression result for TL and TW estimates based on k=0.021 value were given in Figure 11 and Figure 12, respectively. When looking at the ridge regression results in relation to TL, except for the HH and ED variables, the coefficients for the other predictor variables were found to be significant (p<0.05). The effects of predictor variables on TL were determined as HL>PR>PD>BH, from largest to smallest. The regression equation was estimated as follows:

TL= -0.96 + 0.79PR + 0.70PD + 0.45BH + 1.67HL
(Adj-R-square= 0.9726)

```
Coefficients: for Ridge parameter K= 0.021
           Estimate Estimate (Sc) StdErr (Sc) t-value (Sc) Pr(>|t|)
Intercept   -0.9739   -28860.2432   2010.7854     -14.3527   <2e-16 ***
PR           0.7901      177.3161     19.2664       9.2034   <2e-16 ***
PD           0.6966      144.3384     19.3184       7.4715   <2e-16 ***
HH           0.3102       25.0948     20.3016       1.2361   0.2187
BH           0.4484       54.3799     17.2127       3.1593   0.0020 **
HL           1.6704      177.2640     19.8192       8.9441   <2e-16 ***
ED           0.5447        6.3969     12.2769       0.5210   0.6033
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Ridge Summary
      R2    adj-R2  DF ridge         F       AIC       BIC
  0.9737    0.9726    4.0019 1315.1340  491.5358 1135.7909
Ridge minimum MSE= 17028.55 at K= 0.021
P-value for F-test ( 4.0019 , 125.0206 ) = 4.210836e-101
```

Figure 11. Results of Ridge regression in relation to TL according to parameter k=0.021

According to the ridge regression results in relation to TW, except for PD and ED variables, regression coefficients for other predictor variables were found to be significant (p<0.05). The effects of predictor variables on TW were determined as BH>HL>HH>PR from largest to smallest. The regression equation for the relationship was estimated as follows:

TW= -143.99+0.76PR+2.43HH+2.91BH+1.95HL
(Adj-R-square= 0.868)

```
Coefficients: for Ridge parameter K= 0.021
           Estimate Estimate (Sc) StdErr (Sc) t-value (Sc) Pr(>|t|)
Intercept -143.9605   -31550.0963   7546.8960      -4.1805   0.0001 ***
PR           0.7636      171.3529     72.3109       2.3697   0.0193 *
PD          -0.3423      -70.9362     72.5063      -0.9783   0.3298
HH           2.4353      197.0011     76.1963       2.5854   0.0109 *
BH           2.9181      353.8624     64.6031       5.4775   <2e-16 ***
HL           1.9522      207.1725     74.3857       2.7851   0.0062 **
ED          -7.6056      -89.3220     46.0779      -1.9385   0.0548 .
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Ridge Summary
      R2    adj-R2  DF ridge        F      AIC       BIC
  0.8739    0.8688    4.0019 167.5603 835.4152 1479.6703
Ridge minimum MSE= 61053 at K= 0.021
P-value for F-test ( 4.0019 , 125.0206 ) = 3.102247e-49
```

Figure 12. Results of Ridge regression in relation to TW according to parameter k=0.021

4. Discussion

For over three decades, ridge regression has proven to be a valuable tool for applied statisticians and should be routinely investigated in the context of collinear multiple regression (McDonald, 2009). As a regularized regression method, ridge regression reduces the chances of overestimating coefficients, and makes predictive and explanatory models more informative and interpretable, especially for datasets with strong collinearity. The parameter estimation also has less variance than that generated by iterating variables and is therefore more stable (Niu et al., 2019). Ridge regression adjusts the least squares estimator by reducing the impact of multicollinearity on parameter estimates. The resulting equation consists of least squares and estimates that are not biased but are also more stable; This is an important stability for models to be used in prediction (Rikardsen & Johansen, 2003). Sahin et al (2018) tried to estimate the impact of some traits on the albumen index value that is one of the interior egg quality traits using the Least Squares (LS), Ridge Regression (RR) and Principal Components Regression (PCR). Due to the multicollinearity between independent variables, the standard errors and VIF values of the partial regression coefficients in the LS method were found to be quite high. They concluded that the use of RR and PCR analysis methods could be more accurate instead of the LS method under multicollinearity problem.

OLS (k=0 case) and RR (k=0.021) regression analysis results for TL and TW predictions are given together in Figure 13 and Figure 14. When looking at the OLS and RR regression results in Figure 13, it is seen that the regression coefficients are different and the importance levels of the BH and PD predictor variables in RR also increase. However, as expected in RR (Ölmez & Yağanoğlu, 2023), there was a slight decrease in the adj-R-square value (from 0.9848 to 0.97).

```
Coefficients: for Ridge parameter K= 0
             Estimate Estimate (Sc) StdErr (Sc) t-value (Sc) Pr(>|t|)
Intercept     -4.0265   -31703.5376   4563.6421      -6.9470   <2e-16 ***
PR             0.9487      212.8959     49.6621       4.2869   <2e-16 ***
PD             0.7481      155.0240     38.3154       4.0460   0.0001 ***
HH            -0.9275      -75.0313     43.1890      -1.7373   0.0848 .
BH             0.4487       54.4057     24.6356       2.2084   0.0291 *
HL             2.2365      237.3432     45.1410       5.2578   <2e-16 ***
ED             0.1902        2.2332     13.5112       0.1653   0.8690
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Ridge Summary
        R2     adj-R2   DF ridge           F         AIC         BIC
   0.98540    0.98480    5.99995 1391.39510  487.13857 1137.12311
Ridge minimum MSE= 8626.852 at K= 0
P-value for F-test ( 5.99995 , 124.0001 ) = 3.530574e-111
-------------------------------------------------------------------


Coefficients: for Ridge parameter K= 0.021
             Estimate Estimate (Sc) StdErr (Sc) t-value (Sc) Pr(>|t|)
Intercept     -0.9739   -28860.2432   2010.7854     -14.3527   <2e-16 ***
PR             0.7901      177.3161     19.2664       9.2034   <2e-16 ***
PD             0.6966      144.3384     19.3184       7.4715   <2e-16 ***
HH             0.3102       25.0948     20.3016       1.2361   0.2187
BH             0.4484       54.3799     17.2127       3.1593   0.0020 **
HL             1.6704      177.2640     19.8192       8.9441   <2e-16 ***
ED             0.5447        6.3969     12.2769       0.5210   0.6033
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Ridge Summary
       R2    adj-R2  DF ridge          F        AIC        BIC
   0.9737    0.9726    4.0019 1315.1340  491.5358 1135.7909
Ridge minimum MSE= 8626.852 at K= 0
P-value for F-test ( 4.0019 , 125.0206 ) = 4.210836e-101
```

Figure 13. OLS (k=0) and RR (k=0.021) regression analysis results for TL prediction

When OLS and RR regression results for TW are compared (Figure 14), it is seen that the regression coefficients of the variables have changed. It was determined that the significance level of the HL variable increased in RR, and although the regression coefficient of the ED variable was significant in OLS (p<0.05), it was found to be insignificant in RR (p>0.05). As in the TL estimation, the adj-R-square value decreases slightly when TW is modeled with RR (from 0.87220 to 0.8685).

```
Coefficients: for Ridge parameter K= 0
             Estimate Estimate (Sc) StdErr (Sc) t-value (Sc) Pr(>|t|)
Intercept   -144.7147   -29911.3869  17566.9195      -1.7027   0.0911 .
PR             0.9802      219.9666    191.1652       1.1507   0.2521
PD            -1.0923     -226.3406    147.4884      -1.5346   0.1274
HH             3.0434      246.1986    166.2483       1.4809   0.1412
BH             3.0943      375.2203     94.8304       3.9568   0.0001 ***
HL             2.4827      263.4704    173.7621       1.5163   0.1320
ED            -9.2464     -108.5916     52.0088      -2.0879   0.0389 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Ridge Summary
        R2     adj-R2   DF ridge           F         AIC         BIC
   0.89080    0.88640    5.99995  168.53505  837.59178 1487.57632
Ridge minimum MSE= 61053 at K= 0.021
P-value for F-test ( 5.99995 , 124.0001 ) = 3.830558e-57
-------------------------------------------------------------------


Coefficients: for Ridge parameter K= 0.021
             Estimate Estimate (Sc) StdErr (Sc) t-value (Sc) Pr(>|t|)
Intercept   -143.9605   -31550.0963   7546.8960      -4.1805   0.0001 ***
PR             0.7636      171.3529     72.3109       2.3697   0.0193 *
PD            -0.3423      -70.9362     72.5063      -0.9783   0.3298
HH             2.4353      197.0011     76.1963       2.5854   0.0109 *
BH             2.9181      353.8624     64.6031       5.4775   <2e-16 ***
HL             1.9522      207.1725     74.3857       2.7851   0.0062 **
ED            -7.6056      -89.3220     46.0779      -1.9385   0.0548 .
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Ridge Summary
       R2    adj-R2  DF ridge          F        AIC        BIC
   0.8739    0.8688    4.0019  167.5603  835.4152 1479.6703
Ridge minimum MSE= 61053 at K= 0.021
P-value for F-test ( 4.0019 , 125.0206 ) = 3.102247e-49
```

Figure 14. OLS (k=0) and RR (k=0.021) regression analysis results for TW prediction

Among some measure on collinearity, it was pointed that the VIF (individual measure) had a better performance on the fitted model on the morphology of *C. macropomum*. Also, the presence of collinearity indicates that a substantial part of the information in one or more of these covariates is redundant (Rivas Villegas et al., 2024). Therefore, the higher adj-R-square values in OLS compared to RR could be expected. As supported in our case, the relationships of both TL and TW with the predictor variables, a small decrease was observed in the adjusted R-square values obtained by ridge estimation compared to the adjusted R-square values obtained by the OLS estimation method. However, a significant decrease in VIF values outweighs this drawback, indicating that the models are more stable. In the study, estimates were made using all fish caught, without distinguishing between male and female individuals. In a purely ecological study, this situation can be considered a limiting situation in terms of the generalizability of the study results. However, this can be ignored since the current study was not an ecological study and the focus of the study was how ridge regression was used methodologically in evaluating the fish morphology-growth relationship.

5. Conclusion

Fish growth occurs in length and weight. Although it is known that there is a relationship between morphometric characteristics fish growth, modeling is limited to simple linear regression due to the high level of multicollinearity among predictive variables. One of the methods used in cases of multicollinearity is the ridge regression method. However, in this approach, determining the optimum ridge regression parameter (penalization term, k or lambda) can often be speculative and difficult. In the study, using multiple linear models for total length (TL) and total weight (TW) ($Y = \alpha + \beta_1 X_1 + ... + \beta_n X_n + \varepsilon$), the holistic effect of metric features on these two features was estimated using ridge regression. Measurements of a total of 126 *Capoeta damascina* individuals sampled from the Euphrates River (around Adıyaman province, Türkiye) were used as the data set. Predorsal length (PR), head length (HL), post dorsal length (PD), head height (HH) and eye diameter (ED) predictive metric features were considered. The ridge optimum k-parameter, which produces acceptable values (VIF<10) of the variance inflation factor (VIF) of the predictor variables, which are indicators of multicollinearity in the modeling, was determined as k=0.021 in both models. With the Ridge estimation method, in the relationship between total height (TL) and predictive metric traits ($TL = -0.96 + 0.79PR + 0.70PD + 0.45BH + 1.67HL$), except for ED and HH, other predictive variables were statistically significant (p<0.05) (Adj-R-square=0.97). Except for eye diameter and post dorsal length (PD), it was determined that the effects of other metric traits on total weight were significant (p<0.05), and the relationship between total length and important predictive features was estimated as $TW = -143.99 + 0.76PR + 2.43HH + 2.91BH + 1.95HL$

(Adj-*R*-square=0.87). The methodological approach and findings in this study may contribute to filling a gap in the literature regarding the relationships between fish growth and morphology. Additionally, it could enhance better growth predictions for different fish species, aiding sustainable fisheries management and the selective cultivation of desired fish traits in aquaculture by improving the understanding of morphometric features.

Acknowledgements

This research was supported by the Adıyaman University Scientific Research Projects (BAP) Coordination Unit, Project No. 2012/001, and was carried out with the legal consent of the Ministry of Food, Agriculture, and Livestock, General Directorate of Fisheries and Aquaculture (permission date 05.04.2012; permission number: 01515). Thank you to the relevant institutions**.**

Funding Information

This work was supported by Adıyaman University Scientific Research Projects (BAP) Coordination Unit, Project Number: 2012/001.

Conflict of Interest

The authors declare that they have no known competing financial or non-financial, professional, or personal conflicts that could have appeared to influence the work reported in this paper.

Ethical Statement

The paper is not currently being considered for publication elsewhere, and it reflects the authors' own research and analysis in a truthful and complete manner.

References

Adeboye, N. O., Fagoyinbo, I. S., & Olatayo, T. O. (2014). Estimation of the Effect of Multicollinearity on the Standard Error for Regression Coefficients. *Journal of Mathematics*, *10*(4), 16-20. https://doi.org/10.9790/5728-10411620

Albayrak, A. S. (2005). Çoklu doğrusal bağlanti halinde enküçük kareler tekniğinin alternatifi yanlı tahmin teknikleri ve bir uygulama. *Uluslararası Yönetim İktisat ve İşletme Dergisi*, *1*(1), 105-126. (in Turkish)

Alp, A., Akyüz, A., Kara, C., & Güneş, H. (2013). *Assessment of movements of Capoeta capoeta in Bulam Stream, Euphrates River, and efficiency of the fish passage* [Oral presentation]. International Symposium on Fisheries and Aquatic Sciences, Erzurum, Turkey.

Alp, A., Kara, C., Büyükçapar, H. M., & Bülbül, O. (2005). Age, Growth and Condition of *Capoeta capoeta angorae* Hanko 1924 from the Upper Water Systems of the River Ceyhan, Turkey. *Turkish Journal of Veterinary & Animal Sciences*, *29*(3), 665-676.

Arechavala-Lopez, P., Sanchez-Jerez, P., Bayle-Sempere, J. T., Sfakianakis, D. G., & Somarakis, S. (2012). Morphological Differences between Wild and Farmed Mediterranean Fish. *Hydrobiologia*, *679*, 217-231.

Asadollah S., Soofiani, N.M., Keivany, Y., & Hatami, R. (2017). Age and Growth of the Mesopotamian barb, *Capoeta damascina*, in Central Iran. *Iranian Journal of Fisheries Sciences*, *16*(2), 511-521. https://doi.org/10.22092/ijfs.2018.114675

Büyükuysal, M. Ç., & Öz, İ. İ. (2016). Çoklu Doğrusal Bağıntı Varlığında En Küçük Karelere Alternatif Yaklaşım: Ridge Regresyon. *Düzce Üniversitesi Sağlık Bilimleri Enstitüsü Dergisi*, *6*(2), 110-114. (in Turkish)

Clabaut, C., Bunje, P. M., Salzburger, W., & Meyer, A. (2007). Geometric Morphometric Analyses Provide Evidence for the Adaptive Character of the Tanganyikan Cichlid Fish Radiations. *Evolution*, *61*(3), 560-578. https://doi.org/10.1111/j.1558-5646.2007.00045.x

Daoud, J. I. (2017). Multicollinearity and Regression Analysis. *Journal of Physics: Conference Series*, *949*, 012009. https://doi.org/10.1088/1742-6596/949/1/012009

Forrester, J. B., & Kalivas, J. H. (2004). Ridge Regression Optimization Using a Harmonious Approach. *Journal of Chemometrics*, *18*(7-8), 372-384. https://doi.org/10.1002/cem.883

Gao, Y., Fu, Z., Yang, J., Yu, M., & Wang, W. (2024). Spatial–temporal Differentiation and Influencing Factors of Marine Fishery Carbon Emission Efficiency in China. *Environment, Development and Sustainability*, *26*(1), 453-478. https://doi.org/10.1007/s10668-022-02716-6

Golub, G. H., Heath, M., & Wahba, G. (1979). Generalized Cross-Validation as a Method for Choosing a Good Ridge Parameter. *Technometrics*, *21*(2), 215-223. https://doi.org/10.1080/00401706.1979.10489751

Karakaş, S. (2008). *Çoklu doğrusal bağlantı problemi ve yanlı regresyon tahmincileri* [Master's thesis, İstanbul Universty]. (in Turkish)

Katz, M. H. (2006). *Multivariable analysis, A practical guide for clinicians* (2nd ed.). Cambridge University Press.

Keppeler, F. W., Montaña, C. G., & Winemiller, K. O. (2020). The Relationship between Trophic Level and Body Size in Fishes Depends on Functional Traits. *Ecological Monographs*, *90*(4), e01415.

https://doi.org/10.1002/ecm.1415

Khalaf, G., & Shukur, G. (2005). Choosing Ridge Parameter for Regression Problems. *Communications in Statistics - Theory and Methods*, *34*(5), 1177-1182. https://doi.org/10.1081/STA-200056836

Kyritsi, S., & Moutopoulos, D. K. (2018). Mouth Size and Body Length Relations for Freshwater Fish Species. *Turkish Journal of Fisheries and Aquatic Sciences*, *18*(11), 1331-1332. https://doi.org/10.4194/1303-2712-v18_11_10

Lenhardt, M., Jarić, I., Cvijanović, G., Kolarević, J., Gačić, Z., Smederevac-Lalić, M., & Višnjić-Jeftić, Ž. (2012). Comparison of Morphological Characters between Wild and Cultured Sterlet (*Acipenser ruthenus* L.). *Slovenian Veterinary Research*, *49*(4), 177-84.

Liyandja, T. L., Armbruster, J. W., Poopola, M. O., & Stiassny, M. L. (2022). Evolutionary Convergence in Body Shape Obscures Taxonomic Diversity in Species of the African *Labeo forskalii* Group: Case Study of *L. parvus* Boulenger 1902 and *L. ogunensis* Boulenger 1910. *Journal of Fish Biology*, *101*(4), 898-913. https://doi.org/10.1111/jfb.15148

Mami, A. M., Abdalla, A., & Bin Ismaeil, E. H. (2021). Selecting the Optimal Value of Penalty Parameter K in Ridge Regression Estimators. *International Journal of Sciences: Basic and Applied Research*, *60*(5), 189-209.

Martins, B. O., Franco-Belussi, L., Siqueira, M. S., Fernandes, C. E., & Provete, D. B. (2021). The Evolution of Red Blood Cell Shape in Fishes. *Journal of Evolutionary Biology*, *34*(3), 537-548. https://doi.org/10.1111/jeb.13757

McDonald, G. C. (2009). Ridge Regression. Wiley *Interdisciplinary Reviews: Computational Statistics*, *1*(1), 93-100. https://doi.org/10.1002/wics.14

Mojekwu, T. O., & Anumudu, C. I. (2015). Advanced Techniques for Morphometric Analysis in Fish. *Journal of Aquaculture Research & Development*, *6*(8), 1-6. https://doi.org/10.4172/2155-9546.1000354

Nandi S., & Saikia S.K. (2015). Role of Morphometry in Determining the Feeding Success of Small Freshwater Fish Species: Multivariate Analysis of *Amblypharyngodon mola*, *Puntius ticto*, and *Esomus danricus*. *Archives of Polish Fisheries*, *23*, 79-90. https://doi.org/10.1515/aopf-2015-0009

Niu, S. Q., Haake, D. M., & Knouft, J. H. (2019). Gut Content Biomass of Stream Fishes: Associations with Streamflow Velocity, Population Density, and Prey Abundance at Two Spatial Scales. *Hydrobiologia*, *826*(1), 85-97. https://doi.org/10.1007/s10750-018-3719-4

Okamura, H., Yamashita, Y., & Ichinokawa, M. (2017). Ridge Virtual Population Analysis to Reduce the Instability of Fishing Mortalities in the Terminal Year. *ICES Journal of Marine Science*, *74*(9), 2427-2436. https://doi.org/10.1093/icesjms/fsx089

Oke, J., Akinkunmi, W. B., & Etebefia, S. O. (2019). Use of Correlation, Tolerance and Variance Inflation Factor for Multicollinearity Test. *GSJ*, *7*(5), 652-659.

Ölmez, P., & Yağanoğlu, A. M. (2023). Use of Ridge Regression Analysis in the Multicollinearity in Animal Science. *Journal of Biometry Studies*, *3*(2), 32-38. https://doi.org/10.61326/jofbs.v3i2.02

Orbach, D. N., Donovan, M., & Purchase, C. F. (2019). Sexually Selected Traits are Larger and More Variable in Male than Female Beach-spawning Capelin, *Mallotus villosus*. *Journal of Fish Biology*, *95*(6), 1385-1390. https://doi.org/10.1111/jfb.14145

Özkale, M. R., & Altuner, H. (2021). Bootstrap Selection of Ridge Regularization Parameter: A Comparative Study via a Simulation Study. *Communications in Statistics - Simulation and Computation*, *52*(8), 3820-3838. https://doi.org/10.1080/03610918.2021.1948574

Pino, M. M., Rivas, D. V., Carrera, C. O., Avila, N. C., Salazar, T. M., Díaz Merino, H., Gastañaduy, C. C., Ysa, R. S., Lozado, J. C., Santos, R. C., Gastañaduy, D. C., & Palacios, L. V. (2021). Lasso Model for Morphological Covariation Patterns between *Colossoma macropomum* and *Piaractus orinoquensis* × *Colossoma macropomum* Hybrid. *Journal of Biochemical Technology*, *12*(1), 45-48. https://doi.org/10.51847/pFILL2lk3G

Revelle, W., & Revelle, M. W. (2015). Package 'psych'. *The comprehensive R archive network*, 337(338).

Rikardsen, A. H., & Johansen, M. (2003). A Morphometric Method for Estimation of Total Lipid Level in Live Arctic Charr: A Case Study of Its Application on Wild Fish. *Journal of Fish Biology*, *62*(3), 724-734. https://doi.org/10.1046/j.1095-8649.2003.00062.x

Rivas Villegas, D., Rivero Alzamora, C., Carrera Osorio, C., Calderón Ramírez, L., Rojas Correa, L., Cáceres, N., Sánchez Palacios, J., Del Carpio Franco, C., Vásquez Grados, M., Salinas Cruz, L., Rojas Ponce, K., Figueroa Rodríguez, J. J., & Quipas Bellizza, M. M. (2024). Comparison of Collinearity Indices for Linear Models in Agricultural Trials. *OnLine Journal of Biological Sciences*, *24*(2), 195-207. https://doi.org/10.3844/ojbsci.2024.195.207

Sahin, M., Yavuz, E., & Uckardes, F. (2018). Multicollinearity Problem and Bias Estimates in Japanese Quail. *Pakistan Journal of Zoology*, *50*(2), 757-761. http://dx.doi.org/10.17582/journal.pjz/2018.50.2.757.761

Thompson, C. G., Kim, R. S., Aloe, A. M., & Becker, B. J. (2017). Extracting the Variance Inflation Factor and other Multicollinearity Diagnostics from Typical Regression Results. *Basic and Applied Social Psychology*, *39*(2), 81-90. https://doi.org/10.1080/01973533.2016.1277529

Turan, C. (1999). A Note on the Examination of Morphometric Differentiation among Fish Populations: The Truss System. *Turkish Journal of Zoology*, *23*(3), 259-264.

Ullah, M. I., Aslam, M., & Altaf, S. (2018). lmridge: A Comprehensive R Package for Ridge Regression. *The R Journal*, *10*(2), 326-346.

Williams, G. J. (2011). *Data mining with Rattle and R: The art of excavating data for knowledge discovery*. Springer.

Winans, G. A. (1987). *Using morphometric and meristic characters for identifying stocks of fish* [Oral presentation]. The Slock Identification Workshop, Florida, USA.

Yazıcı, R. & Yazıcıoğlu, O. (2020). Some Morphometric Relationships of Wels Catfish (*Silurus glanis* L., 1758) Inhabiting Sıddıklı Dam Lake (Kırşehir, Turkey). *Journal of Anatolian Environmental and Animal Sciences*, *5*(2), 199-204. https://doi.org/10.35229/jaes.718647