Classifying five ornamental fish species of Cichlidae family by use of logistic regression and discrimination linear analysis

Mooraki N.[1*]; Omrani M.[1]; Khajehrahimi A.E.[1]; Azhdari P.[2]

Received: February 2020 Accepted: July 2020

Abstract

This experiment was conducted to classify five ornamental fish species *Heros severus, Astronotus ocellatus, Pterophyllum scalare, Amphilophus citrinellus*× *Paraneetroplus melanurus* and *Amphiprion perideraion* regard to the growth indices BWI, SGRW and SGRL by Use of Logistic Regression and Discrimination Linear Analysis, in a completely random design with five treatments (5 species) each with four replicates with a density of 25 specimens per replicate and totally equal to 500 ornamental fish. Feeding trail with Iso-nitrogenous and Iso-caloric diets lasted for 90 days, and every thirty days fish biometry was being practiced. At the end of the trial, data were collected to find the best class for each species regard to the least distances with other species. The results obtained from two statistical models were compared and it was concluded that the Logistic Regression model in species classification based on the type of nutrition and considering the BWI and SGRW indices rather than the describes by Discrimination linear analysis, is more appropriate. But the SGRL index was not considered as an appropriate factor for the separation of fish samples according to species factor in either of these two models.

Keywords: Ornamental fish, Cichlid, Statistical models, Discrimination linear analysis, Logistic regression

1-Department of Fisheries Science, College of Marine Science and Technology, North Tehran Branch, Islamic Azad University (IAU), Tehran, Iran.

2-Department of Statistics, College of Science, North Tehran Branch, Islamic Azad University (IAU), Tehran, Iran.

*Corresponding author's Email: Nargess_Mooraki@Yahoo.com

Introduction

Regard to the rapid progression of rearing Ornamental fishes, increasing effort are made for optimizing the culture condition. One of the entanglements now a day is deficiency of natural resources, especially water and energy, the two key factors for a rearing plant. Therefore any economize approach which can be effective in saving natural resources, would be appreciated by the industry. One of the strategies which could be effective for saving the resources; is integrated culture of species, with similar demands for ambient condition, accordingly evaluating the effectiveness of Synchronic culture of similar species, could be a guide for practitioners to have an economic activity. In this regard classification of similar species due to the growth indices with statistical point of view will be an explainable pathway for selecting and rearing species both for experts and amateurs in this progressive domain. The aim of the present study was to classify five ornamental fish species belong to Cichlid family through the study of growth pattern by use of Logistic Regression and Discrimination Linear Analysis.

Material and methods

In the present trial, 500 specimens belong to five species Severum (*Heros severus*), Oscar (*Astronotus ocellatus*), Angel (*Pterophyllum scalare*), Paret (*Amphiprion penderaion*) divided to five treatment each with four replicates were reared for 90 days, with 30 days of interval for biometry. The specimens were fed with a similar iso-nitrogenous and iso-caloric diet (Table 1).

Table1: The composition of diet used for feeding fish species.

Composition of diet	(%)
Dry Matter	97.89
Crude Protein	41.48
Fiber	3.20
Crude Lipid	8.80
Ash	19.00
Carbohydrate	24.9

To maintain the water quality, bio-filters and a central oxygen pump and heater were used in aquaria, 50 percent of water in each aquarium replaced every other three days at the beginning and every other five days at the end of the trial. Temperature, dissolved oxygen, pH, nitrite were measured daily (Table 2).

Table 2: The Physic-chemical parameters of the aquaria measured during the trial.

Temperature	Nitrite (mg/L)	Dissolved Oxygen (mg/L)	pH
27±1°C	0.02	6.5±0.5	7.3±0.5

During the rearing period, all the groups were fed three times a day (6, 12, and 18) at the base of 4 percent of their biomass and also with respect to their eagerness and appetite. For evaluating the growth, biometry was done once every 15 days. The weight was measured by use of digital balance (bearing 0. 01g) and the lengths were

obtained by use of caliper; 12 hours before conducting biometry feeding of fish were cut to reduce and avoid stress and oxygen depletion. Growth performance regard to the obtained data was measured by calculating the equations as bellow:

Specific growth rate (%); SGR= [ln final body weight _ ln initial body weight / days of experiment] ×100
Specific growth rate (%); SGR= [ln final body length - ln initial body length / days of experiment] ×100
Weight gain = (Final weight - initial weight) / (initial weight) ×100

Statistical methods

At the end of the rearing period, In order to find the best class for each species as to have the least distance with other four species, tow statistical model including Discrimination Linear Analysis (DLA) (Bendel and Afifi, 1979; Johnson and Wichern, 2007) and Logistic Regression (Lr) (James *et al*., 2013) were used. The obtained results of models were compared.

Results and discussion

Due to evaluate and compare the growth rate of five species of Cichlid family; the hypothesis of analysis was being focused as mentioned below:

H.1: Multicollinearity shouldn't be observed between factors; as the presence of multicollinearity shows that at least two of the factors have a strong correlation.

H.2: There should be a linear correlation between the factors.

H.3: There should not be outliers for factors.

In order to analyze the above hypothesis, the normality of SGRw, SGRL, and BWI data were evaluated at p=0.000 by use of Kolmogorov-Smirnov and Shapiro-Wilk tests.

Regard to Figs. 1a and 1b there is a linear relation for BWI, but there are outliers which rejected H.2.

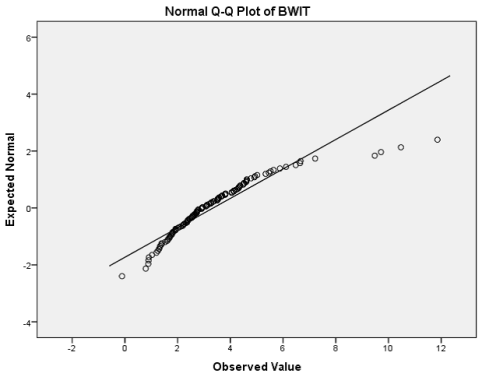

Figure1a: Linear Correlation of BWI.

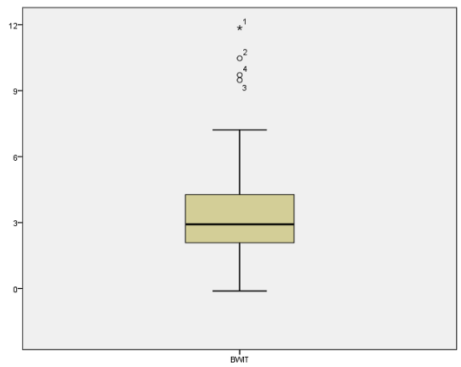

Figure 1b: Outliers of BWI

The same results were obtained for SGRw and SGRL (Figs. 2a, 2b, 3a, and 3b).

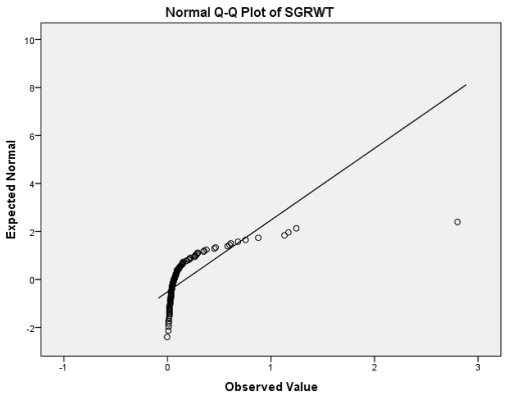

Figure 2a: Linear Correlation of SGRw.

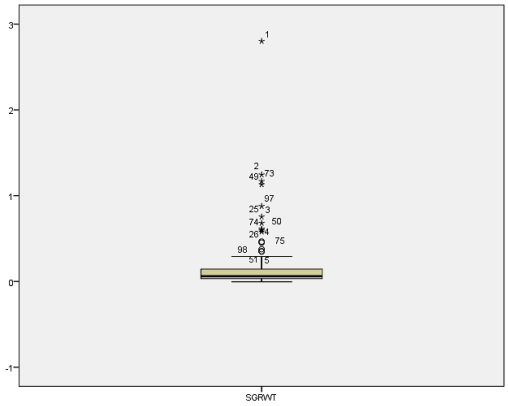

Figure2b: Outliers of SGRw.

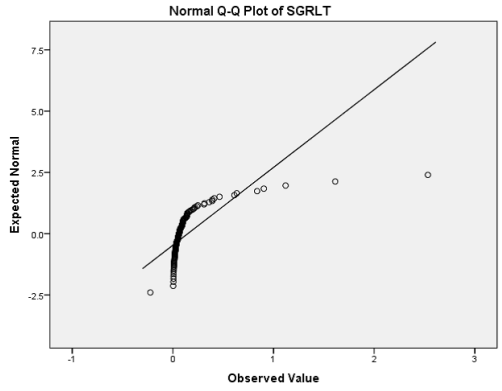

Figure 3a: Linear Correlation of SGRL.

In the next stage, testing H.2, the presence of correlation between factors was checked by calculating Pearson coefficient. As it is shown in Table 3 there is a strong multicollinearity between BWI, SGRw and SGRl.

Regard to the hypothesis H.1, H02, and H03 and also none proper classification of species as depicted by matrix in Figs. 4a, 4b, 4c, 4d, 4e; and then it are obvious that Discriminant Analysis is not an appropriate model for classifying the specimens regard to growth indices.

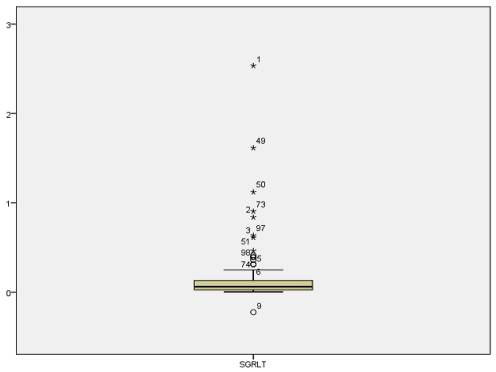

Figure 3b: Outliers of SGRL.

Logistic regression analysis

Due to the flexibility of this analytical procedure and no necessity of introducing constraints, compare to the Discriminant Analysis, the pathway was run by presenting the factors separately.

Regard to the results presented in Table 4, BWI index can be an effective factor for classification of the 5 species, when Paret considered as the reference, with probability equal to 0.012. The same concept can expressed the usefulness of SGRw as an effective index for classification the 5 species regard to Table 5.

Table 3: Pearson correlation coefficient between factors.

	BWI	SGRw	SGRl
Pearson Correlation	1	0.677	0.590
Sig.(2-tailed)	-	0.000	0.000
N	120	120	120
Pearson Correlation	0.677	1	0.923
Sig.(2-tailed)	0.000	-	0.000
N	120	120	120
Pearson Correlation	0.590	0.928	1
Sig.(2-tailed)	0.000	0.000	-
N	120	120	120

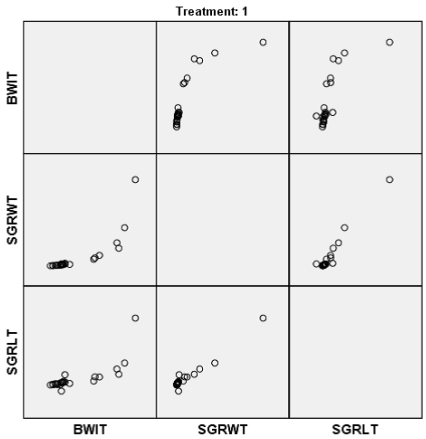

Figure 4a: Classification matrix of *P. scalare* regarding to the growth indices.

Regard to the obtained results from Discriminant Linear Analysis and also the H03 hypothesis, it is obvious that biomass, specific growth rate of weight and length could not be introduced as appropriate factors for classifying the selected 5 species, as it is depicted a linear correlation for biomass but the outliers were affected the matrix significantly, so the use of Discriminant analysis is not recommended for classification the species regard to growth performance indices.

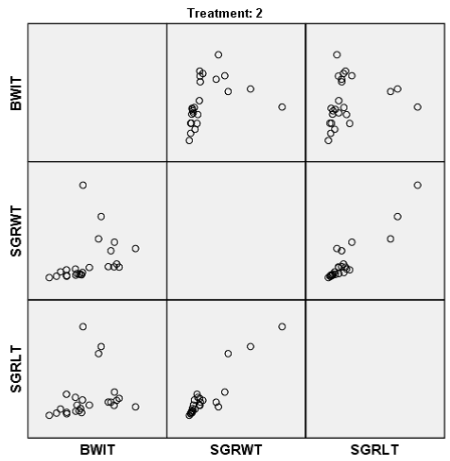

Figure 4b: Classification matrix of *A. citinellus* * *P. melanurus* regarding to the growth indices.

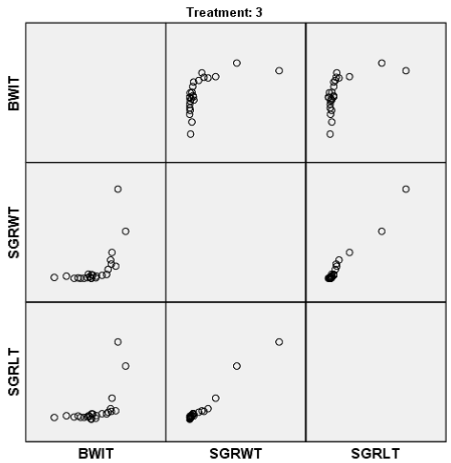

Figure 4c: Classification matrix of *A. ocellatus* regarding to the growth indices.

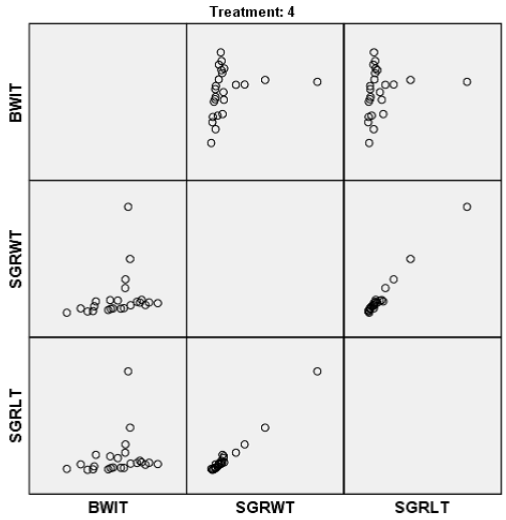

Figure 4d: Classification matrix of *C. severum* regarding to the growth indices.

Table 4: Analyzing the regression model by use of goodness of fit regarding to the BWI.

Model	Model Mining Criteria	Likelihood Ratio Tests		
	-2logLikelihood	Chi-square	df	Sig.
Intercept Only	3334-379	26.325	4	0.000
Final	353-554			
Pearson		460.344	463	0.584

Table 5: Analyzing the regression model by use of goodness of fit regarding to the SGRw.

Model	Model Mining Criteria	Likelihood Ratio Tests		
	-2logLikelihood	Chi-square	df	Sig.
Intercept Only	265-386	3.9777	4	0.409
Final	238-332			
Pearson		479.00	472	0.402

On the other hand, the model obtained from Logistic Regression pathway (Sturdivant *et al*., 2013; Peduzzi *et al.*, 1996) revealed that BWI and SGRw can properly classify the studied species; although SGRL was not an appropriate index for this goal. In conclusion weight gain is a reliable factor for classifying the species and prepares an economical condition for integrated rearing of species with similar growth pattern. And logistic Regression model is more accurate compare to discrimination Analysis for evaluating the growth pattern; the reason which support the explanation, is may be due to abnormal distribution of predictor variable and the effectiveness of some qualitative noise factors in the trial condition.

References

Bendel, R.B. and Afifi, A.A., 1979. Comparison of Stopping Rules in Forward Stepwise Discriminant Analysis. *Journal of the American Statistical Association*, 74:777-785

James, G., Witten, D., Hastie, T., Tibshirani, R., 2013. An Introduction to Statistical Learning with Application in R, Springer. 112 P.

Sturdivant, R.X., Lemeshow, S., Hosmer, D.W. Jr., 2013. Applied Logistic Regression, Third Edition, John Wiley & Sons. 397 P.

Peduzzi, P., Concato J., Kemper, E., Holford, T.R. and Feinstein, A.R., 1996. A simulation study of the number of events per variable in logistic regression analysis. *Journal of Clinical Epidemiology*, 49: 1373-1379.

Johnson, R.A. and Wichern, D.W., 2007. Applied Multivariate Statistical Analysis, Sixth Edition. 794 P.