The correlation between energy levels and biological factors across all seasons for smallmouth bass and lake trout

Ze Qian Wu
1002192254
April 2, 2019

Table of Content

Summary

The study focuses on examining the correlation between two different energy concentrations and other biological indicators of two types of fish, Smallmouth Bass and Lake Trout, from using the data collected throughout a year from a lake in Algonquin Provincial Park in Algonquin, Canada. Through performing linear models and correlation tests, there are two indicators, fork length and rounded weight, showed strong potential correlation with energy 2 (Muscle Protein) while GSI and body condition indicator showed relatively weaker correlation. Energy 1(Liver Triglyceride) does not show strong correlation with any biological indicators with the use of linear models and correlation matrix, however, two variables, Heposomatic Index and body condition indicator, have a significant reading after performing gamma regression on Energy 1.

Introduction

Smallmouth bass and lake trout are two species of fish that both can be found in Algonquin Provincial Park in Algonquin, Ontario. Smallmouth bass is a type of warm-water fish that is medium sized, whereas lake trout is a type of cold-water fish native to Ontario[1]. This study measures two energy levels, muscle protein and liver triglyceride, along with associated biological indicators to approach the study of fish's behavior. As the two types of fish have different adaptation to water, it would be interesting to see if they behave differently throughout seasons. The purpose of this study is to examine if there is any interaction across all seasons between smallmouth bass and lake trout's energy level (muscle protein and liver triglyceride) and biological indicators (fork length, rounded weight, body condition factor, Heposomatic Index (HSI), and Gonado-Somatic Index(GSI)[2]).

To explore the research question further based on the dataset collected by collaborator, Timothy Fernandes, a statistical approach is taken in the data analysis to analyze the correlation between energy concentration and other indicators. Another chronological factor, season is used to determine if there is any trend or different behavioral activities that are significant throughout the year. All the biological factors differ within species as every fish has individual length measurements, weight, body condition, and organ indices.

[1] Smallmouth Bass and Lake Trout. (n.d.). Retrieved from https://www.ontario.ca/page
[2] CF, HIS and GSI of fish 2005. (n.d.). Retrieved from https://www.epd.gov.hk/epd/mobile/english/environmentinhk/water/marine_quality/bio_cf05.html

Methods

Raw data was collected from Algonquin provincial park and two types of fish were specifically caught during the process. The collaborator and his team used various fishing methods including angling with a fishing rod and netting, and all the fish were dissected in the lab afterwards. As the number of fish caught varied between trips, some fish were dissected the day they were caught and some were kept in the lab until time was allotted for dissecting. All the specific biological indictors were recorded before and during the process of dissection, including the length, weight, body condition, liver mass, sex, muscle protein and etc[3]. In order to obtain preliminary data that is representative across all seasons, fish were caught in the following sampling months: March, May, June, August, October and November with a relatively bigger sample size in spring and summer. A total of 32 observations were recorded for each fish, with a sample size of 129 fish. In order to have a more randomized sample, the fishing location changed randomly within the lake as well as the fishing methods.

In this study, 9 out of the 32 observations are used to examine any potential correlation between variables. The species and season of the fish are recorded as either "smallmouth bass(SMB)" or "lake trout(LT)" and paired with one of the four seasons. On behalf of the size, the collaborator uses two different length measurements: one indicates the total length, and the other one indicates the fork length, which is a length measurement in millimeter of fish excluding tips of the tail fin. Based on the nature of the study, the fork length is a more appropriate fish size indicator than total length, and therefore only the measurement of fork length is used. The body condition factor is calculated as a function of body mass and length, representing the health of an individual fish. The organ indices of the fish are indicated with

[3] Full list of biological indicator can be found in the raw dataset (CleanData_scripted)

Heposomatic Index (HSI) and Gonado-Somatic Index (GSI). With the context, HSI is s defined as the ratio of fish's liver weight to its body weight and GSI is the ratio of fish's gonad weight to its body weight. According to all the samples that are collected, each fish is labelled as either "male" or "female" with some of them being "unknown" as no significant evidence suggest their gender. Sex maturity is labelled with four numbers: 10 – Immature (Gonads are not developed), 22 – Partial (Gonads are developing), 23 – Ripe (Gonads are fully developed) and 40 - Spent (Gonads are emptied). Considering the selected biological indicators from the full raw dataset, we are interested in analyzing the data based on a research question: Across all seasons, does any of the indicators correlate with Energy 1(Muscle Protein) and Energy 2(Liver Triglyceride)?

To answer the research question, linear models and correlation matrices are used to determine the potential correlation that exists. Linear models describe the relationship between a response variable (biological indicators) as a function of a predictor variable (energy levels), which allow us to see the relationship between the raw datasets before building any statistical models. The correlation matrices also clearly illustrate the correlation coefficients between variables. As Energy 1 does not appear to be normal, log transformation is performed on Energy 1 to make the dataset conform to normality. After performing linear models and correlation matrices, Energy 1 still does not appear to have any correlation with any other biological indicators. In order to get a valid result for Energy 1, gamma regression is used to scale parameters for exponential-response data. Since the dataset is non-negative, it is appropriate to run gamma regression based on Energy 1.

Result

To examine the correlation between variables, a correlation matrix is used to display the correlation between each variable that we are interested in exploring. Figure 1 below illustrates the correlation between each biological indicator and energy concentrations. By reading the values of r, the fork length and rounded weight have a relatively strong negative correlation of r = - 0.81 and r = - 0.83 with muscle protein. This result shows that these two variables are correlated with fish's muscle protein with rounded weight being slightly more statistically significant. On the other hand, GSI and body condition factor also has a value of r that is 1> r > 0.5 or - 1< r< - 0.5, which shows that they have weaker correlations with Energy 2.

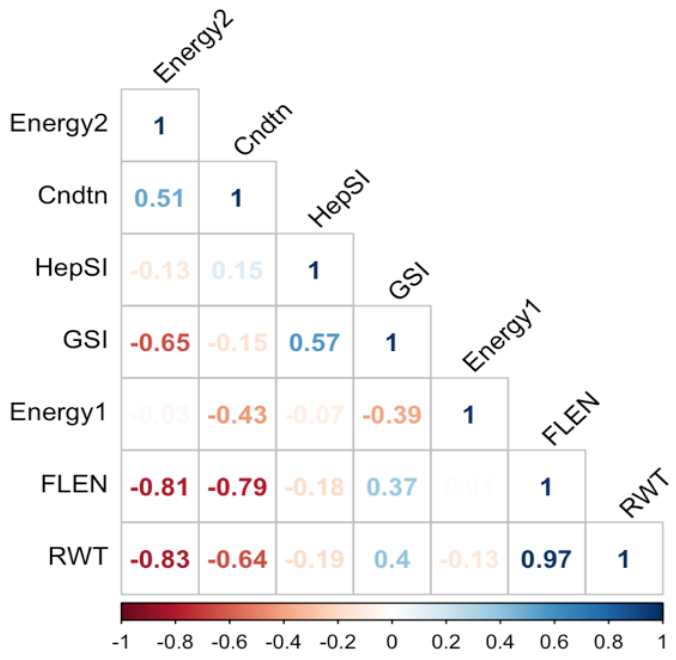

Figure 1. The correlation matrix

As there are two variables that appear to have relatively strong correlation, we can now plot the residuals out and see if the plots are consistent with the correlation matrix across all seasons with both species.

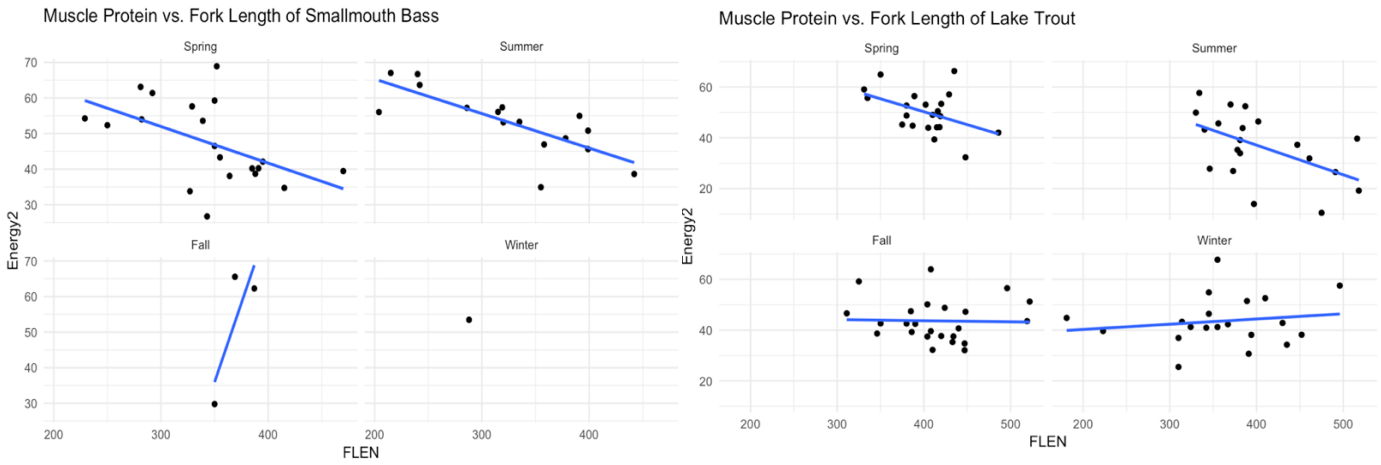

Figure 2. Muscle Protein vs. Fork Length Across All Season with Two Separated Species

According to Figure 2, it is presented that during spring and summer, both of the species follow the same pattern of having less Energy 2 when their fork length is longer. However, the trend in fall and winter are not as representative. The reason why that fall and winter show less of a negative trend might due to the lack of sample and weather factors. In general, both of the species conform a negative correlation between Energy 2 and fork length.

Based on the correlation matrix, little to no evidence shows that there is any correlation between Energy 1 and other biological factors. In order to explore further on seeing if there is any potential correlation, gamma regression is used to extract useful information. From performing gamma regression analysis, two variables, HSI and body condition factor appeared to be significant with HSI being more statistically significant. The graph below highlights the significant p-value of HSI and body condition factor.

Fixed Effects	P-value	Significance Level
HSI	0.000134	***
Body Condition Factor	0.026535	*
Fork Length	0.624155	-
Rounded Weight	0.641177	-

Figure 3. Summary table of Gamma Regression Energy1 vs. Biological Factors

Conclusion and Discussion

Based on the research interest of the study, we are attentive to observe if there exists any correlation between the two energy concentrations of fish and other biological identities. From performing linear models, correlation matrix and gamma regression, there exists a potential strong correlation between Energy 2 with fork length and rounded weight, as well as a week correlation between Energy 2 with GSI and body condition factor. The negative correlation indicates that the longer the fork length/the heavier the body weight, the lower the muscle protein. On the other hand, Energy 1 presents a potential stronger correlation with HSI and a weaker correlation with body condition factor. Therefore, HSI is closely associated with liver triglyceride, which means that changing in HSI of fish will have a direct impact to a great population of fish's liver triglyceride level.

In general, both Energy 1 and Energy 2 indicate that they are correlated with body condition factor on different levels. With the statistical results observed, we can conclude that the fork length and rounded weight of a fish can be used as reliable indictors of fish's muscle protein, and HSI can be used as a reliable indicator of fish's liver triglyceride.

There exists limitation in the fish behavioral study as all of the samples are from the same lake, therefore the result does not represent the two types of fish in general as other geographical factors can directly affect the data and result. Considering that fishing and dissecting across all seasons is a long and repetitive process, there might be bias during the whole process of data collecting and recording. Based on the raw data, the time interval varies between the fish was caught and dissected. The freshness of the fish might also affect some of the biological variables. Besides all the biases that occurred during the process of data collection, one of the most considerable limitation for this study is the sample size. Having a sample size of n=129 is not big enough to conduct any complex statistical analysis that might help with obtaining a valid and reliable result. Insufficient sample size can directly cause biased result and findings. As a way to eliminate the effect of insufficient information, a much bigger sample size is preferred for examining correlation.