Exploring the Impact of Diverse Biological Factors on Metabolic Rates in a Wide Spectrum of Living Organisms through Linear Regression Analysis

Project Topic: This project is exploring different factors that affect metabolic rates including a large set of living organisms. This data set was a compilation of 20+ published sources and looks at different factors including wet mass, dry mass, and temperature to name a few. This data set does include one row on humans within the 10,000+ rows.

Research Question: One research question that will be addressed within this capstone project is: What variables influence metabolic rate of various living organisms?

> **Hypothesis**: **Null hypothesis**- There is no significant influence of any of the studied variables on the metabolic rate of the included living organisms.
> **Alternate Hypothesis**- At least one of the studied variables have a significant influence on the metabolic rate of the included living organisms.

Context: Looking at metabolic rates of different living organisms, can aid with ecological understanding. Metabolic rates are crucial in understanding the flow of an ecosystem. With a better understanding of metabolic rates across living organisms, we can assess different environmental factors like the flow of energy, growth and reproduction, and the impact of other species within the environment. "Assessing the relationship between energy flux and the quantity of biomass it sustains offers the potential to understand the biological "carrying capacity" for ecosystems on Earth and beyond. Our work supports this understanding by quantifying the energy–biomass relationship for the global biosphere and an environmentally diverse range of its components, and by exploring the factors—including the impact of humanity—that affect that relationship." (Hoehler et al., 2023)

Data: Due to the nature of this projects topic, data will not need to be collected and will instead reference a completed data set.

> The data set that will be used is a database that was compiled by multiple different published sources. These sources then compiled data to included 10,000+ rows of various living organisms. The data included shows descriptive factors like group, domain, kingdom, phylum, class, order, family, genus, species, and type of metabolic rate. Other factors include wet mass, dry mass, dry mass to wet mass ratio, carbon mass, temperature, and metabolic rate.

The data was published in the Proceedings of the National Academy of Sciences Journal and was distributed under Creative Commons Attribution-NonCommercial-NoDerivatives License 4.0 (CC BY-NC-ND). This license states that we are allowed to use and adapt the data, but it is necessary to give appropriate credit. The only restrictions lie in commercial use which is not utilized within this personal and education project.

Data Gathering: I will be utilizing a pre-existing data set and will conduct a secondary data analysis on this data set.

Data Analytics Tools and Techniques: I will utilize multiple linear regression to determine influential variables. The model created will then be reduced utilizing backward stepwise elimination to determine the most influential variables on metabolic rate within the

given living organisms. I will also use Variance Inflation Factor to look for multicollinearity and the RSE will be used to compare initial and final models to look at effectiveness.

Justification of Tools/Techniques: Multiple linear regression is an appropriate technique used to analyze the data collected due to the focus being on relationships between the main continuous variable (metabolic rate) and the chosen variables within the data set. Backward stepwise elimination reduces the model based on p-values and helps focus on those variables that have more influence on the metabolic rate. VIF looks for multicollinearity and RSE helps compare the initial and reduced models.

Project Outcomes: The key anticipated project outcomes will include a list of variables and how influential each variable is on metabolic rate. We will also have other outcomes that will help us determine how effective our model is and how our variables influence metabolic rate. We will have coefficient values, p-values, R-squared values, F-test values, and a final regression equation.

Sources:
Hoehler, T. M., Mankel, D. J., Girguis, P. R., McCollom, T. M., Kiang, N. Y., & Jørgensen, B. B. (2023). The metabolic rate of the biosphere and its components. *Proceedings of the National Academy of Sciences of the United States of America*, 120(25), e2303764120. https://doi.org/10.1073/pnas.2303764120