EXPLORING THE IMPACT OF DIVERSE BIOLOGICAL FACTORS ON METABOLIC RATES IN A WIDE SPECTRUM OF LIVING ORGANISMS THROUGH LINEAR REGRESSION ANALYSIS

BY: MAKAYLA AVENDANO

ABOUT ME – BRIEF INTRODUCTION

Makayla Avendano

- **Education:**
  - Masters of Science, Data Analytics → October 2023
    - Western Governors University
  - Bachelors of Science, Biology → May 2021
    - University of New Mexico
- **Professional Background:**
  - Operations Coordinator → FedEx
    - Logistics Analytics

PROBLEM

- Metabolic rate is the energy expended by a living organism to sustain life and basic bodily functions.
- There is still significant opportunity for knowledge advancement and scientific exploration.
- An understanding of metabolic rate and its interactions with various factors is important to understand an organism's ability to exist, survive, and thrive in the world.
- Linear regression model will help assess relationship between various factors and metabolic rate.

HYPOTHESIS

- **Hypothesis:**
  - At least one of the studied variables have a significant influence on the metabolic rate of the included living organisms.

- **Null Hypothesis:**
  - There is no significant influence of any of the studied variables on the metabolic rate of the included living organisms.

DATA ANALYSIS PROCESS:
DATA INFORMATION

- Data was not collected – referenced a completed data set
- Data set was found on the Data is Plural newsletter that highlighted the data set that was created and published within the environmental sciences journal in Proceedings of the National Academy of Sciences.
- Database from multiple sources - 10,000+ rows of various living organisms

Column	Description
Group	Group of Taxa
Domain	Taxon Rank
Kingdom	Taxon Rank
Phylum	Taxon Rank
Class	Taxon Rank
Order	Taxon Rank
Family	Taxon Rank
Genus	Taxon Rank
Species	Taxon Rank
Publication species name	Taken from the original measurement
Tsn	Taxonomic serial number
Tsn rank	Taxonomic rank
Wet mass (g)	Wet mass of organism (in grams)
DM/WM	Dry mass to wet mass ratio
Dry mass (g)	Dry mass of organisms (in grams)
Carbon Mass (g)	Carbon mass of organism (in grams)
Metabolic Rate (W, at T)	Metabolic rate, temperature from study
T (C)	Temperature (in Celsius)
Metabolic Rate (W, at 25C)	Metabolic rate, temperature corrected
Type of Metabolic Rate	Type of metabolic rate
O2 Original J/mL	Conversion of J/mL O2
Reference	Publication data was taken from
Reference Code	Citation
Comments	Any relevant information
Individual or Average Measurement	Whether measurement from study was average or individual

DATA ANALYSIS PROCESS:
DATA CLEANING

- Data cleaning is a necessary step to make sure we are addressing potential data integrity threats, the unnecessary skewing of data, and the potential inflation or deflation of values.
  - Duplicates
    - No duplicates in the data
  - Missing Values
    - 5 numerical variables with missing values
    - All values were replaced with the median values
  - Outliers
    - 5 variables with outliers evident from boxplots
    - All outliers were retained

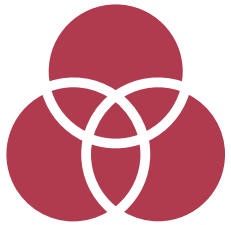
DATA ANALYSIS PROCESS:

DATA ANALYSIS

- Multiple Linear Regression – tool to look at relationship between variables
- Variance Inflation Factor (VIF) - used initially to address multicollinearity within the data
- Backward Stepwise Elimination – method to reduce the model

- Results
  - Model – final reduced model with important model statistics
  - Variables – final variables that are shown to have an affect on the dependent variable

OUTLINE OF FINDINGS



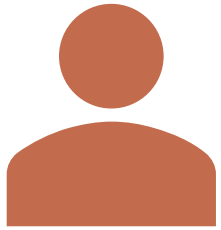
Variable Elimination

Variance Inflation Factor (VIF):

16 variables

Backward Stepwise Elimination:

15 variables



Model

Initial Model RSE:

481.74

Final Model RSE:

481.41

OUTLINE OF FINDINGS

- Reject the null hypothesis
- Final Variables:
  - Carbon Mass
  - Mammal Group
- Prob (F-statistic): 0.0
- F-statistic: 16,510

OLS Regression Results

Dep. Variable:	metabolic_rate	R-squared (uncentered):	0.758
Model:	OLS	Adj. R-squared (uncentered):	0.758
Method:	Least Squares	F-statistic:	1.651e+04
Date:	Tue, 17 Oct 2023	Prob (F-statistic):	0.00
Time:	12:56:41	Log-Likelihood:	-79974.
No. Observations:	10529	AIC:	1.600e+05
Df Residuals:	10527	BIC:	1.600e+05
Df Model:	2		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
carbon_mass	0.0017	9.73e-06	179.676	0.000	0.002	0.002
Group_Mammal	105.1998	11.382	9.243	0.000	82.890	127.510

Omnibus:	28264.895	Durbin-Watson:	0.395
Prob(Omnibus):	0.000	Jarque-Bera (JB):	758597328.191
Skew:	32.683	Prob(JB):	0.00
Kurtosis:	1316.350	Cond. No.	1.18e+06

OUTLINE OF FINDINGS

- Coefficients indicate the effect that each variable has on metabolic rate.
- Living organism classified as being a part of the mammal group, metabolic rate increases by 105.1998.
- One unit increase in carbon mass is associated with a 0.017 increase in metabolic rate.

VARIABLE	COEFFICIENT VALUE
Mammal Group	105.1998
Carbon Mass	0.017

LIMITATIONS OF TECHNIQUES AND TOOLS

Missing Value - Imputation
Influence the data and skew it unnecessarily

Outlier - Retention
Influence the data and skew it unnecessarily

Multiple Linear Regression
Highly sensitive to outliers

Variance Inflation Factor (VIF)
May not eliminate multicollinearity entirely

Backward Stepwise Elimination
Could cause the elimination of variables that have causal effects on dependent variable

PROPOSED ACTIONS

More exploratory approach on most influential variables: Carbon Mass and the Mammal Group

Performing exploratory analysis

Creating visualizations to identify trends

Looking at different statistical breakdowns can reveal a more in-depth interpretation of the influence of various factors on metabolic rate

<u>Future Study:</u>

ANOVA (exploratory analysis)

Closer look at endotherms

EXPECTED BENEFITS OF THIS STUDY

- Carbon Mass is more impactful then dry and wet mass
  - Use carbon mass instead of wet/dry mass for further study on metabolic rate
- Mammals increase metabolic rate by 105
  - Use this information to address metabolic plasticity in terms of environmental changes
- Biologists and scientists benefit from further understanding of metabolic rate of living organism
- Areas of advancement and exploration include:
  - Flow of energy throughout an ecosystem
  - Energy to biomass relationship within the biosphere
  - Metabolic plasticity