Seneca POLYTECHNIC

# Statistics for Analytics

BAN100NAA - PROFESSOR: SAMANEH GHOLAMI

Maaz Hussain
STUDENT ID: 173714221

# ASSIGNMENT 3:

# CORRELATION

## Problem 1

# Analysis of Fitness Data

## Introduction

This analysis examines the relationships between key fitness variables, including oxygen consumption (oxy), runtime, resting pulse (rstpulse), running pulse (runpulse), and maximum pulse (maxpulse). Scatter plots and correlation coefficients were used to explore the nature of these relationships, and the limitations of correlation analysis were considered.

**The CORR Procedure**

| 5 Variables: | oxy runtime rstpulse runpulse maxpulse |
|---|---|

**Simple Statistics**

| Variable | N | Mean | Std Dev | Sum | Minimum | Maximum |
|---|---|---|---|---|---|---|
| oxy | 31 | 47.37581 | 5.32723 | 1469 | 37.38800 | 60.05500 |
| runtime | 31 | 10.58613 | 1.38741 | 328.17000 | 8.17000 | 14.03000 |
| rstpulse | 31 | 53.45161 | 7.61944 | 1657 | 40.00000 | 70.00000 |
| runpulse | 31 | 169.64516 | 10.25199 | 5259 | 146.00000 | 186.00000 |
| maxpulse | 31 | 173.77419 | 9.16410 | 5387 | 155.00000 | 192.00000 |

## Descriptive Statistics

The simple statistics summarize the characteristics of the dataset:

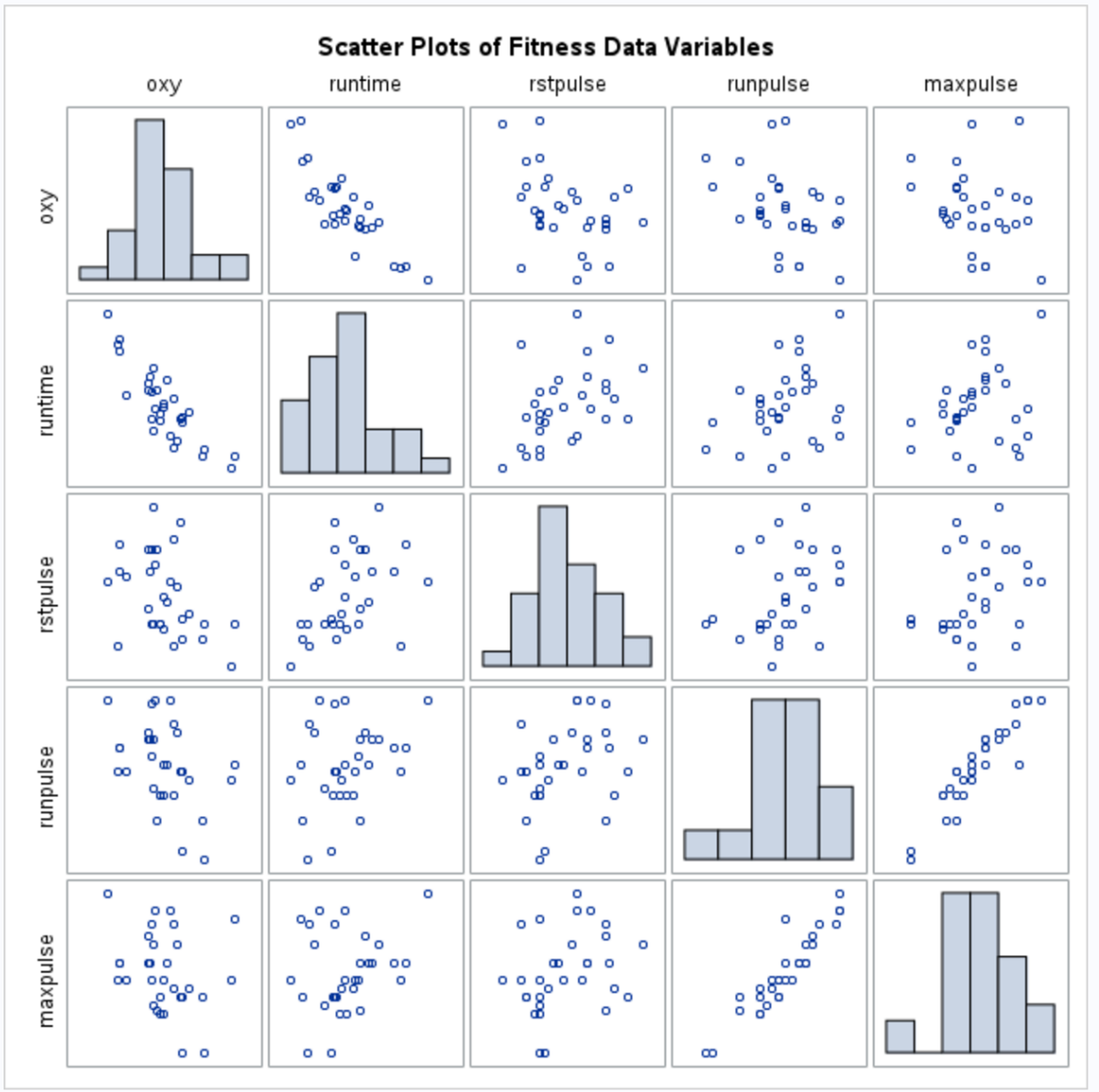
- **Oxygen Consumption (Oxy)**: The mean value is 47.38 ml/kg/min, with a range from 37.39 to 60.06, indicating a moderate spread.
- **Runtime**: Average runtime is 10.59 minutes, with a low variability (SD: 1.39), suggesting participants' times are generally close to the mean.
- **Resting Pulse (Rstpulse)**: Pulse rates at rest averaged 53.45 bpm, ranging from 40 to 70 bpm.
- **Running Pulse (Runpulse) and Maximum Pulse (Maxpulse)**: These showed higher mean values (169.65 bpm and 173.77 bpm), consistent with exertion during activity.

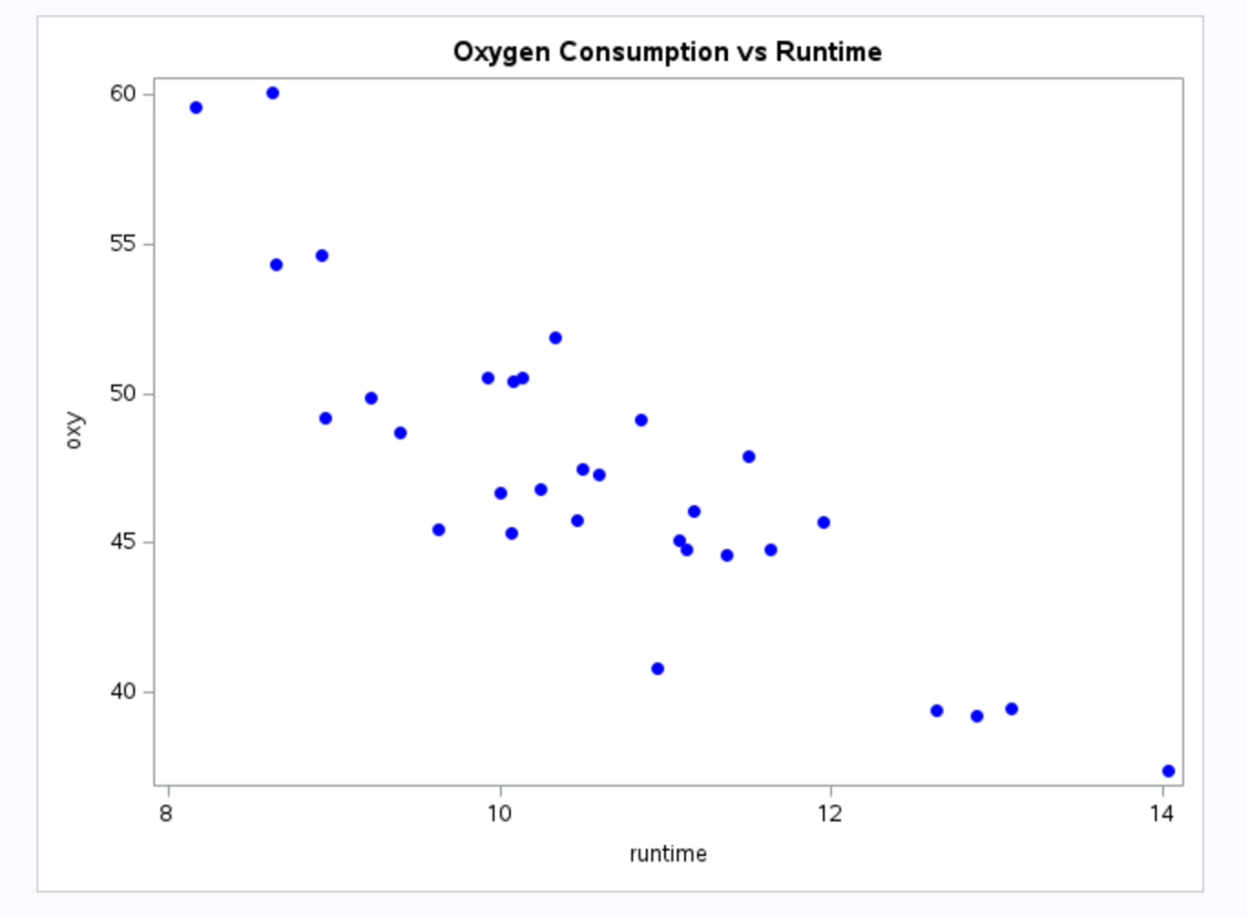
| Pearson Correlation Coefficients, N = 31 | | | | | |
|---|---|---|---|---|---|
| | oxy | runtime | rstpulse | runpulse | maxpulse |
| oxy | 1.00000 | -0.86219 | -0.39936 | -0.39797 | -0.23674 |
| runtime | -0.86219 | 1.00000 | 0.45038 | 0.31365 | 0.22610 |
| rstpulse | -0.39936 | 0.45038 | 1.00000 | 0.35246 | 0.30512 |
| runpulse | -0.39797 | 0.31365 | 0.35246 | 1.00000 | 0.92975 |
| maxpulse | -0.23674 | 0.22610 | 0.30512 | 0.92975 | 1.00000 |

## Correlation Analysis

The Pearson correlation matrix reveals several important relationships:

- **Oxygen Consumption vs Runtime**:
  - Correlation: -0.862 (strong negative). Faster runners tend to have higher oxygen consumption, consistent with improved cardiovascular fitness.
- **Oxygen Consumption vs Resting Pulse**:
  - Correlation: -0.399 (moderate negative). Participants with higher resting pulses tend to consume less oxygen, suggesting lower fitness levels.
- **Running Pulse vs Maximum Pulse**:
  - Correlation: 0.930 (very strong positive). Indicates these variables are tightly linked, as expected during physical exertion.
- **Runtime and Resting Pulse**:
  - Correlation: 0.450 (moderate positive). Longer runtimes are slightly associated with higher resting pulses.


Scatter Plots of Fitness Data Variables
oxy
runtime
rstpulse
runpulse
maxpulse
oxy
runtime
rstpulse
runpulse
maxpulse

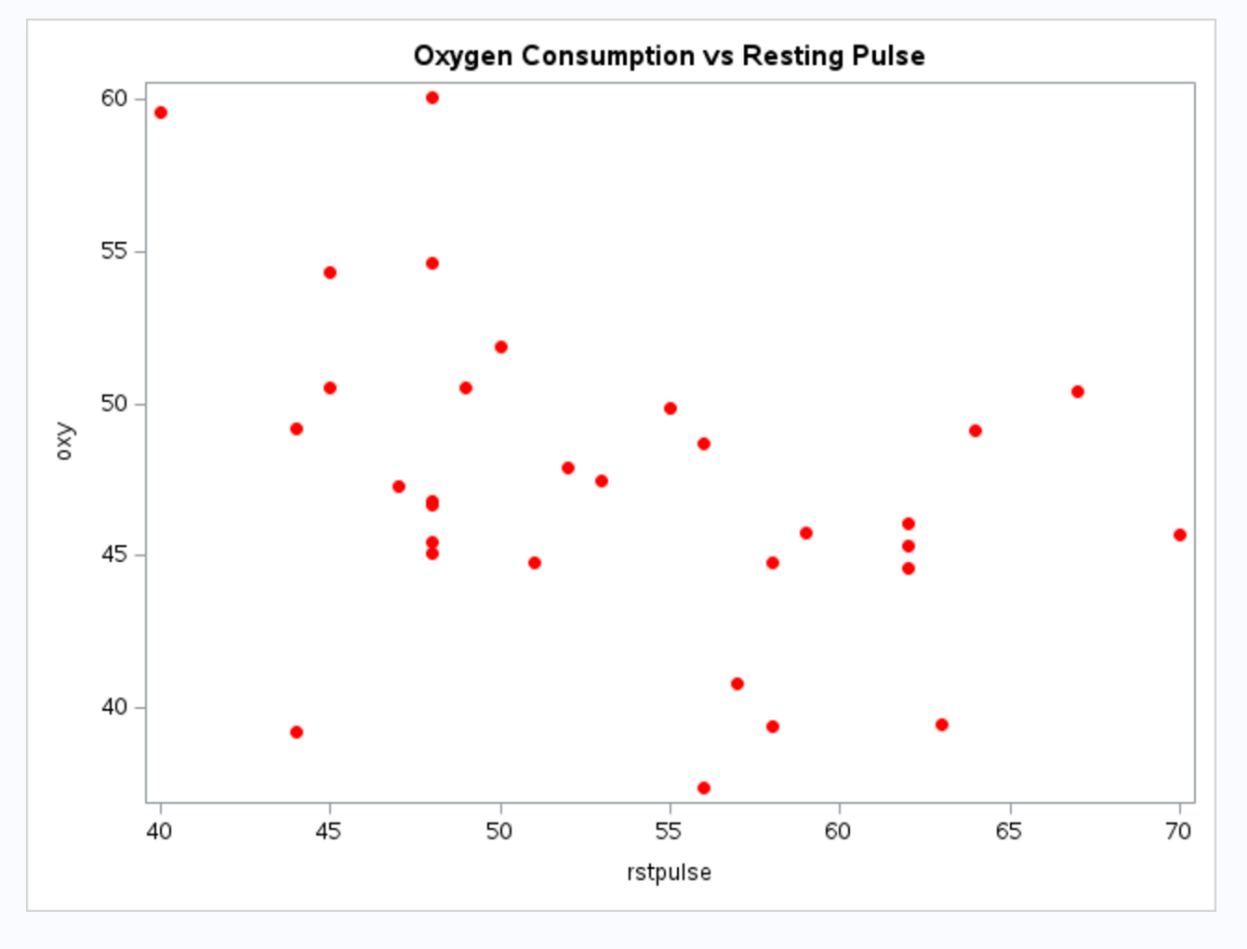



## Scatter Plot Interpretation (Oxygen Consumption vs Runtime):

A strong negative trend is evident. Shorter runtimes are associated with higher oxygen consumption, indicative of better cardiovascular efficiency.

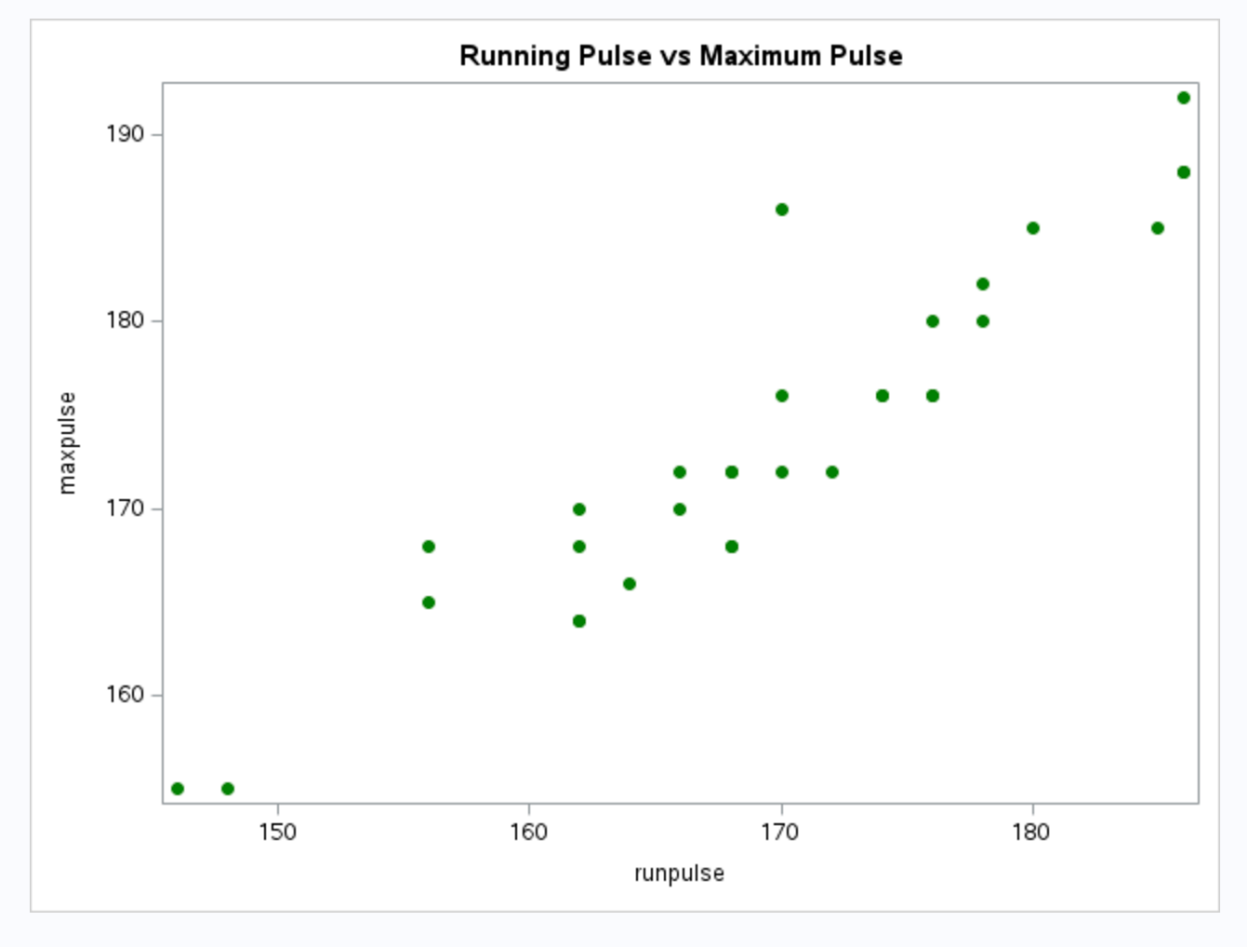
### *Outliers*

The scatter plot of **Oxygen Consumption vs Runtime** revealed an outlier with a runtime of ~14 minutes and relatively low oxygen consumption. This could suggest unique physiological traits or measurement errors, requiring further investigation. Other plots showed consistent patterns with no major anomalies.



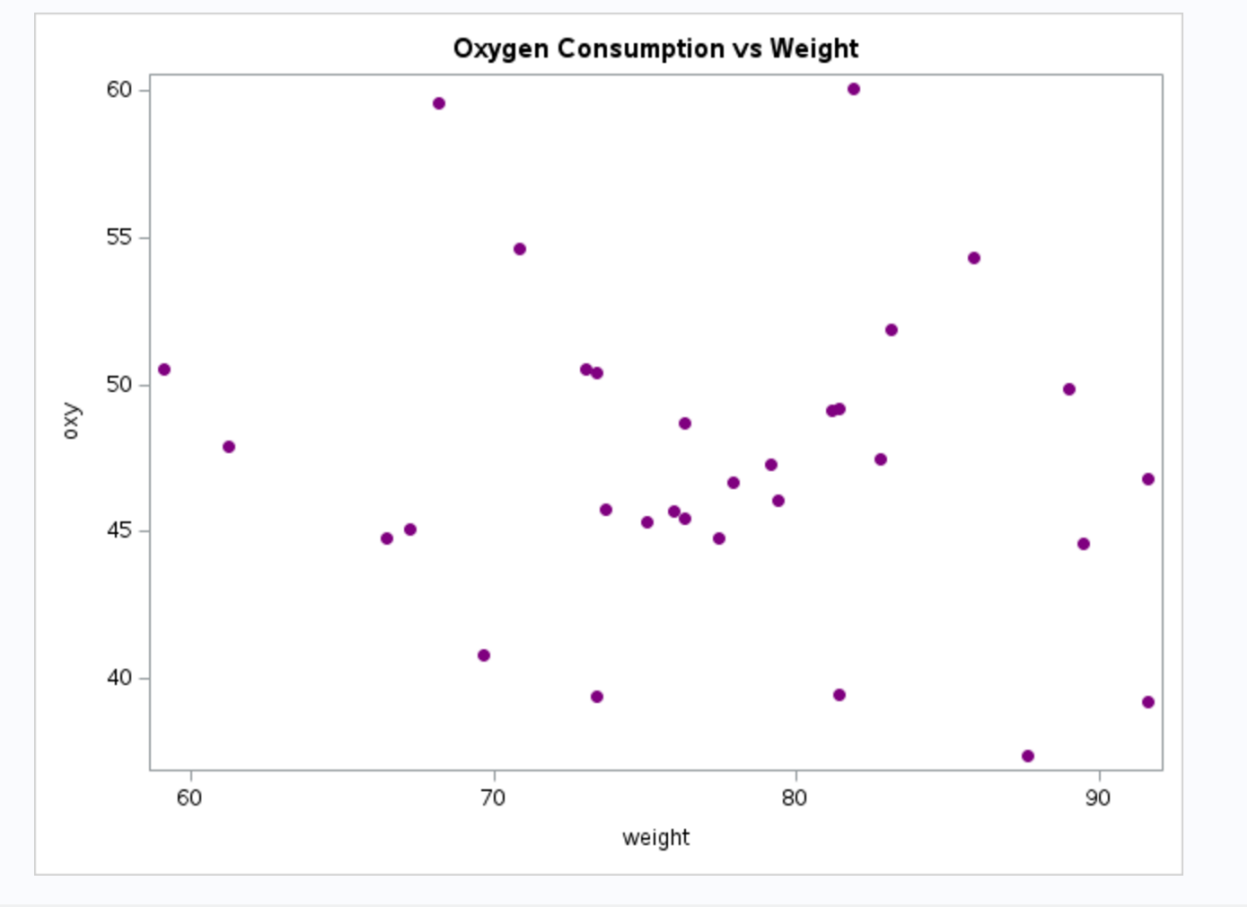

## Oxygen Consumption vs Resting Pulse:

A moderate downward trend is observed, aligning with the correlation coefficient. Fitness levels may decrease as resting pulse increases.




## Running Pulse vs Maximum Pulse:

The plot shows a nearly perfect positive linear relationship, demonstrating the close physiological link between these two variables during exertion.




## Oxygen Consumption vs Weight:

The relationship appears weak and possibly non-linear, suggesting weight has an indirect or minimal impact on oxygen efficiency.

## Limitations of Correlation Coefficient

While the correlation coefficients provide valuable insights, they have several limitations:

- **Causality**: Correlation does not imply causation. For instance, while runtime and oxygen consumption are strongly related, other variables like weight, age, or training intensity could influence both.
- **Non-linear Relationships**: The correlation coefficient only captures linear associations. Non-linear patterns, if present, may not be adequately reflected.
- **Outliers**: Extreme values in oxygen consumption or pulse rates could skew the results, potentially overstating or understating relationships.
- **Multicollinearity**: High intercorrelation between variables (e.g., running and maximum pulse) may pose challenges in regression analysis.

## Conclusion

The analysis highlights the strong interdependence of fitness metrics, particularly the inverse relationship between oxygen consumption and runtime. While the results align with expectations for fitness physiology, the limitations of correlation analysis must be considered. Future analyses could incorporate regression modeling to explore the combined effects of variables like age, weight, and pulse rates on oxygen consumption. Additionally, stratifying the analysis by weight categories may uncover potential trends.