Maximizing Caloric Expenditure: A Comparative Study of Cardio Exercises

Erik M. Painter

Cleveland State University

Abstract

The demand for effective workout plans continues to grow among those into fitness who seek to optimize their exercise performance and maximize caloric expenditure. This study aims to investigate the effects that different types of cardio exercises have on caloric expenditure. The experiment employs a replicated Latin Square design with 4 replicates: two subjects where each is subjected to both AM and PM sessions. Participants perform three types of cardio exercises (Walking Uphill, Stationary Bike, and Elliptical) at three different durations (10, 20, and 30 minutes) while performing the cardio exercises at three different intensity levels (Low, Medium, and High). The primary research question revolves around determining whether different types of cardio exercises elicit differential effects on caloric expenditure across various durations and intensity levels.

The experiment is conducted with two participants, me, and a friend (due to logistical constraints), where the conducting of the experiment will be done by randomly applying treatment factors and blocking, ensuring fairness and reliability of the findings. During each cardio session, caloric expenditure is measured using an Apple Watch, Fitbit, or the cardio machines built in calorie counter. All these mechanisms calculate calories burned based on weight, height, and movement of the individual. The Apple Watch and Fitbit utilize personal health information, such as heart rate and metabolic rates of the user to track caloric expenditure more precisely. The accuracy of the devices' caloric expenditure calculations is based on scientific principles, particularly the metabolic rate, which is the rate at which the body expends calories. Individual metabolic rates are predicted using factors such as age, gender, weight, and height, ensuring precise measurement of calorie expenditure during workouts.

Preliminary results suggest there may exist a difference in caloric expenditure based on the type of cardio exercise performed. The findings of this study have significant implications for fitness enthusiasts, trainers, and even casual fitness persons, as this study will offer evidence-based guidance on the selection of cardio exercises to maximize their workout performance and caloric expenditure. Furthermore, this research contributes to the broader body of knowledge on exercise physiology and exercise strategies.

1. Introduction

The motivation for this project comes from the overall increase in demand for workout plans that maximize or optimize caloric expenditure during a workout session, but also from my own curiosity. With a growing focus on optimizing exercise performance and maximizing caloric expenditure, there is a need to understand how different types of cardio exercises impact calorie burning. This study aims to address this need by investigating the effects of various cardio exercises on caloric expenditure.

The research question guiding this study is: "Do different types of cardio workouts elicit differential effects on caloric expenditure across various durations and intensity levels?" This question aligns with the scope of inference provided by the dataset, which allows for comparisons of caloric expenditure based on different types of cardio workouts, durations, and intensity levels.

The target population for this study includes individuals interested in fitness and exercise, particularly those seeking to optimize their workout routines for enhanced caloric expenditure and exercise performance.

The variables in this study include:

- Response variable: Caloric expenditure during cardio workouts (Cal. Burned)
- Explanatory variable:
    - Cardio Exercise
        - Walking Uphill
        - Stationary Bike
        - Elliptical
- Blocking Variables:
    - Duration
        - 10 min
        - 20 min
        - 30 min
    - Intensity Level
        - Low
        - Medium
        - High

The experiment employs a replicated Latin Square design with 4 replicates. Two subjects, each participating in both morning (AM) and evening (PM) sessions, undergo different combinations of cardio exercises, durations, and intensity levels. Randomization ensures the fair allocation of treatment factors and blocking to minimize any potential biases.

A Latin Square was created with three factors: Type of Cardio (A, B, C), Duration (10, 20, 30), and Intensity Level (Low, Medium, High), with Duration and Intensity Level designated as blocking factors. Standard Latin Square matrices were randomly selected, and rows, columns, and makeup of the square were randomized using random procedures and generated numbers. Treatments were then assigned randomly within the Latin Square matrix based on descending order of randomly generated numbers. Once the Latin Square was designed, the order in which the participants performed the 9 different factor combinations for each replicated Latin Square was randomized using R code.

The cause-effect plot below helps visualize the factors influencing caloric expenditure during cardio exercises. These factors include the type of cardio exercise, duration of workout, intensity level, individual metabolic rates, and other factors such as measurement mechanism, location, Machine Brand, etc. Each factor contributes to the overall effect on caloric expenditure.

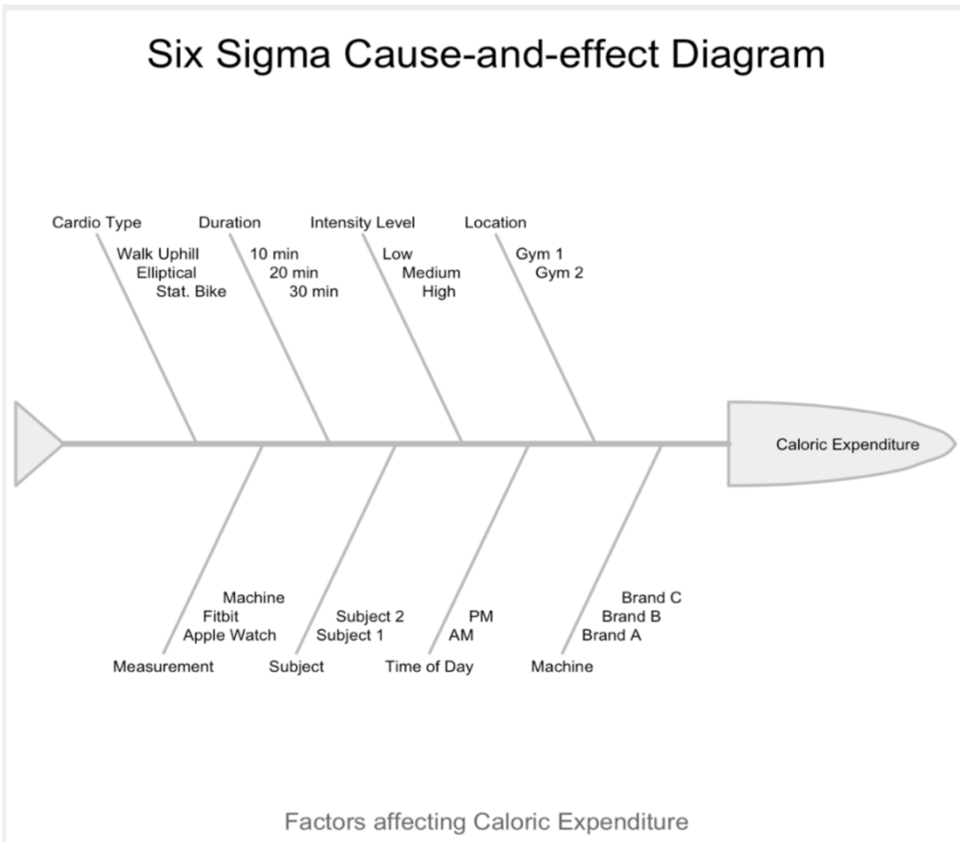

Figure 1: Cause-Effect Plot Illustrating Factors Influencing Caloric Expenditure during Cardio Workouts

The sample for this study consists of two participants, the researcher, and a friend, due to logistical constraints. This is a convenient sample, as the two subjects were not randomly selected from a population. While the small sample size limits generalizability, the rigorous design and randomization procedures aim to ensure the reliability of the findings within this specific sample.

To ensure the statistical robustness of our study, it was imperative to calculate an appropriate sample size. For this experiment, we decided to calculate a sample size based on the power curve for a One-Way ANOVA. This calculation allows us to determine the number of replicates needed to detect meaningful differences in caloric expenditure across various treatments with a desired level of statistical power and significance.

Firstly, the number of treatment levels (k) in our experimental design was determined to be 3. This accounts for the different types of cardio exercises under investigation. Additionally, a significance level (α) of 0.1 and a desired power (1-β) of 0.80 were chosen to control for Type I and Type II errors, respectively. Furthermore, the calculation required an estimate of the standard deviation (σ) of caloric expenditure. This was derived from pilot data collected prior to the main study. The effect size (d) was also determined, representing the magnitude of difference in mean caloric expenditure between treatments that we deemed to be significant. For our study, an effect size of 50 was chosen, indicating that a treatment resulting in 50 extra calories burned compared to another would be considered significant. Using these parameters, we calculated the sample size necessary to detect the specified effect size with the desired level of power and significance.

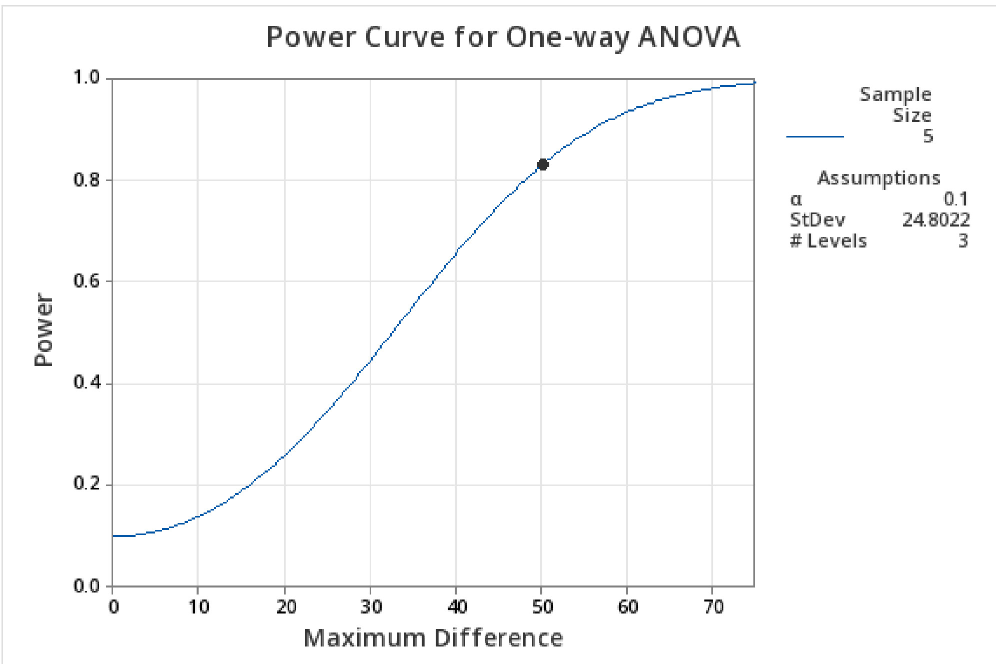

Figure 2: Power Curve based on One-way ANOVA for Sample Size Calculation

Overall, the scope of the analysis is focused on comparing caloric expenditure across different types of cardio exercises, durations, and intensity levels within the context of the study's experimental design and sample size limitations. The analysis and inference are confined to the participants involved, namely me and the other subject, as the sample was selected through convenience sampling rather than random sampling from a broader population.

2. Exploratory Data Analysis

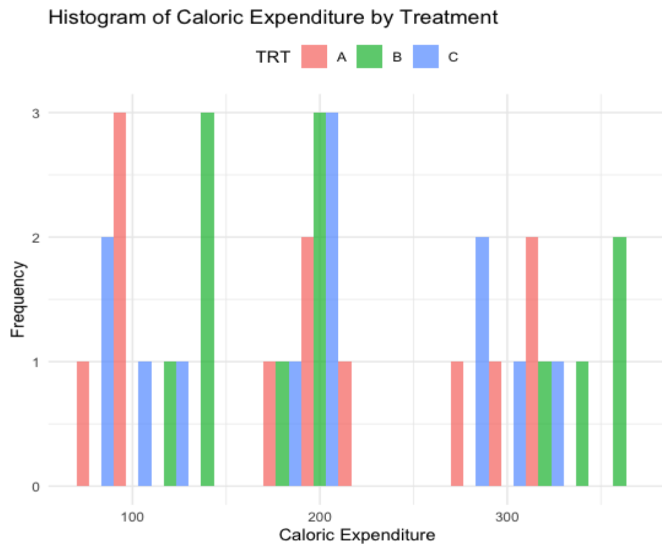

Figure 3: Histogram of Cal. Burned by Treatment

Before running any analysis, we must do some exploratory data analysis to understand the main characteristics of the data and visualize its distribution to gain insights inform our analysis. Summary statistics for caloric expenditure (Cal. Burned) were calculated by treatment level (TRT). Treatment Level A exhibited a mean caloric expenditure of 200.6 calories, ranging from 86.0 to 324.0 calories. Treatment Level B showed a slightly higher mean of 223.2 calories, with caloric expenditure ranging from 120.0 to 355.0 calories. Treatment Level C had a mean caloric expenditure of 195.7 calories, with values ranging from 83.0 to 320.0 calories (Appendix – Figure 8). These findings indicate varying levels of caloric expenditure across different treatment levels. This will guide our hypothesis of is there a difference in treatment effects on our response.

In addition to the summary statistics, the boxplots also reveal important insights into the variability of caloric expenditure within each treatment group (Appendix – Figure 9). Treatment Levels A and B exhibit slightly larger interquartile ranges (IQRs) compared to Treatment Level C, suggesting some variability in caloric expenditure within these groups. This variability may indicate differences in individual responses to the treatments or other factors influencing caloric expenditure. Although no outliers were observed in the data, the boxplot for Treatment Level B shows a higher maximum and higher thresholds for the IQR, indicating potential increases in caloric expenditure compared to other treatments. The larger IQRs for Treatment Levels A and B suggest a wider spread of data points and potentially more diverse responses to the treatments.

Upon examination of the histogram depicting caloric expenditure across all treatment levels, there is evidence of a positive skew. This skewness suggests that the distribution of caloric expenditure is asymmetrical, with a tail extending towards higher values. Such a distribution indicates that there are factor combinations which exhibit higher caloric expenditure compared to the majority of factor combinations. Possible explanations for this positive skewness include variations in metabolic rates, how the caloric expenditure was measured, base health levels, and many more. The presence of positive skewness has implications for the choice of central tendency measure in subsequent analyses, as results based on mean may overestimate the typical caloric expenditure. The skewness could also violate the assumption of normality for certain statistical tests. Further investigation into the underlying factors contributing to this positive skewness could provide valuable insights into individual differences in exercise performance and caloric expenditure.

3. Model & Assumptions

The study employs a Replicated Latin Square Design with the same rows and columns across replications, as well as the same treatment combination layout. The model used in this design is an additive linear effect model. Due to the layout of the Latin Square Design, the model is orthogonal, meaning that the treatment effects and treatment means are orthogonal to the row and column effects. The model used to analyze the data is represented as:

$$Y_{hijk} = \mu + \delta_h + \alpha_i + \beta_j + \tau_k + \epsilon_{hijk}$$

$$h = 1,2,3,4\,;\, i = 1,2,3\,;\; j = 1,2,3\,;\; k = 1,2,3$$

- Y_{hijk} is the observation from the h replicate, i row, j column from the k treatment
- μ is the overall mean
- δ_h is the replicate effect [Fixed]
- α_i is the row effect [Fixed]
- β_j is the column effect [Fixed]
- τ_k is the treatment effect [Fixed]
- ϵ_{hijk} is the random error (E [ϵ_{hijk}] = 0, V [ϵ_{hijk}] = σ^2)

The assumptions of our model are:

1. The errors are normally and independently distributed with mean zero
2. The errors have constant variance σ^2
3. There is no interaction between the blocks and treatments

To estimate treatment effects, we calculate the deviation of caloric expenditure for each type of cardio exercise (walking uphill, elliptical, stationary bike) from the overall mean caloric expenditure. This approach allows us to quantify the impact of each treatment on caloric expenditure relative to the average effect across all treatments. Additionally, we analyze the effects of blocking factors by examining how variations in workout duration (10 minutes, 20 minutes, 30 minutes) and intensity (low, medium, high) influence caloric expenditure. While we are primarily interested in treatment effects, analyzing blocking factor levels can provide valuable insights into certain aspects of the experiment. To assess the significance of treatment effects and identify differences between specific treatment levels, we employ Analysis of Variance (ANOVA). Furthermore, we will conduct post-hoc tests, such as Tukey's HSD test, to determine pairwise differences in mean caloric expenditure between treatment levels. These estimation methods enable us to comprehensively analyze the effects of different treatments and blocking factors on caloric expenditure.

4. Analysis

The ANOVA indicates that the treatment factors (type of cardio exercises) significantly influence caloric expenditure. Specifically, the *TRT* factor has a p-value less than 0.05, indicating statistical significance.

```
Analysis of Variance Table

Response: data$`Cal. Burned`
          Df Sum Sq Mean Sq  F value    Pr(>F)    
Duration   2 261152  130576 707.8110 < 2.2e-16 ***
Intensity  2   2340    1170   6.3426  0.005709 ** 
TRT        2   5195    2598  14.0807 7.189e-05 ***
Replicate  3   1471     490   2.6583  0.069254 .  
Residuals 26   4796     184                       
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Figure 4: ANOVA Results

Since our blocking factors were not randomized, it's important to note that the F-statistic for each blocking factor may not provide a completely accurate assessment. This limitation arises because the F-statistic relies on the assumption of two Chi-Squared distributed variables, which does not hold in this context. However, upon examining both blocking factors, *Duration*, and *Intensity*, it becomes evident that their Mean Square statistics significantly exceed the Mean Square Error (MSE). This observation suggests that our blocking approach effectively mitigated a substantial portion of variation, aligning with our primary objective of implementing blocking.

We now want to analyze the residuals to ensure that the underlying assumptions of the model are met, thus validating the reliability of the statistical inferences drawn from the model.

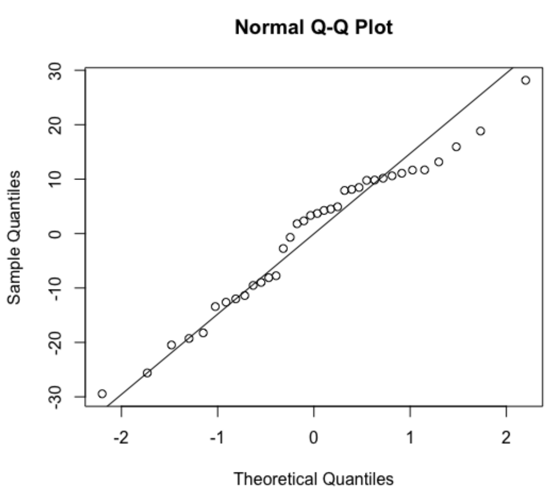

Figure 6: Q-Q Plot demonstrating adherence of residuals to normality assumption

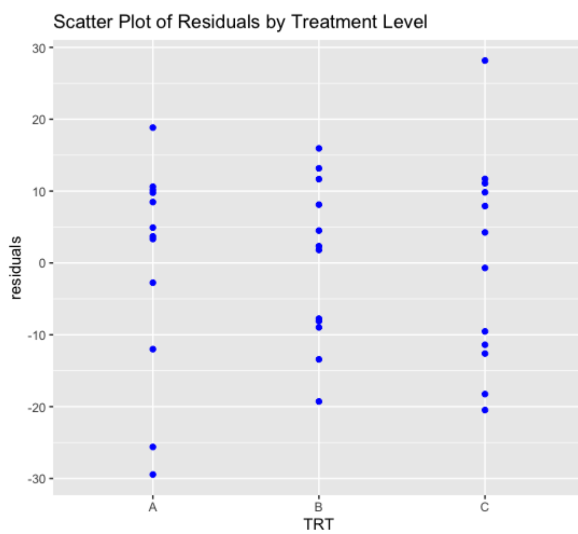

Figure 5: Scatter plot demonstrating adherence of homogeneity of variance assumption

We can see from the *Normal Q-Q Plot* that are residuals are approximately normally distributed. I wanted this assumption to be more concrete, therefore I ran the Shapiro-Wilks test and obtained a p-value of .2741, indicating that the residuals are normally distributed. The *Scatter Plot of Residuals by Treatment Level* shows that our variances are also approximately equal across treatment levels, however, I again wanted to make sure of this, so I ran Levene's test and obtained a p-value of .7147, indicating that there is homogeneity of variance. Lastly, I checked the assumption of no interaction between the blocks and treatment by comparing the residuals and the fitted values for our model and there was no evidence of a pattern or curvilinear shape, which indicates no interaction (Appendix – Figure 10). When all assumptions of a statistical model are met, it signifies that the model accurately represents the data, ensuring the validity, reliability, interpretability, and generalizability of the results.

Given the significant effect of the *TRT* on caloric expenditure (*Cal. Burned*) observed in the ANOVA, conducting Tukey's HSD test will help us to determine which treatment (*TRT*) level yields the highest mean caloric expenditure. This post-hoc analysis will identify the treatment level(s) that differ significantly from each other in terms of their effects on caloric expenditure. Tukey's HSD test was chosen for its ability to control for the family-wise error rate, which safeguards against Type I errors and provides a more conservative approach to multiple comparisons, ensuring reliable identification of significant differences among treatment levels.

```
  Tukey multiple comparisons of means
    95% family-wise confidence level

Fit: aov(formula = lm_model)

$TRT
          diff        lwr        upr     p adj
B-A  22.666667   8.888062  36.445272 0.0010486
C-A  -4.916667 -18.695272   8.861938 0.6533254
C-B -27.583333 -41.361938 -13.804728 0.0001036
```

Figure 7: Tukey's HSD results

Tukey's HSD test was conducted to assess pairwise differences in mean caloric expenditure among the three treatment levels (*TRT*). Treatment Level B exhibited a significantly higher mean caloric expenditure compared to Treatment Level A (mean difference = 22.67 cal., 95% CI [8.88, 36.45], p = 0.001). Conversely, Treatment Level C showed a significantly lower mean caloric expenditure compared to Treatment Level B (mean difference = -27.58 calories, 95% CI [-41.36, -13.80], p = 0.0001). However, there was no significant difference observed between Treatment Level C and Treatment Level A (mean difference = -4.92 calories, 95% CI [-18.70, 8.86], p = 0.653). The comparisons between Treatment Level B and Treatment Levels A and C were statistically significant which means that the observed differences in treatment means is large enough such that it is unlikely to have come from sampling variability or measurement error.

Overall, based on the results, Treatment Level B is associated with a significantly higher mean caloric expenditure compared to Treatment Level A and Treatment Level B. Therefore, Treatment B appears to be the most effective at maximizing caloric expenditure among the three treatment levels.

5. Conclusions & Discussion

In conclusion, this study is an introduction and begins to understand the factors influencing caloric expenditure during cardio exercise, yet it is crucial to recognize its inherent limitations. The scope of the study was constrained by logistical challenges and the inability to obtain a random sample from a population, limiting the generalizability of findings beyond the two subjects examined. Furthermore, the study's statistical power was influenced by the modest sample size. In determining the required sample size for a Latin Square Design, two methods are commonly used. Initially, we opted for a power curve analysis in the context of a One-Way ANOVA. However, this approach operates under the assumption that the total sum of squares comprises solely the Sum of Squares Treatment and the Sum of Squares Error. Consequently, our sample size estimation tends to be inflated, as it assumes a greater need for replicates to account for variability primarily driven by randomness (SSE).

Our total variation in this experiment encompasses not only treatment effects but also those arising from blocking factors and the replicate effect. An alternative method involves using computer simulations to derive an appropriate sample size. Although this calculation offers potential benefits, its intricacies inhibited the ability to accurately perform and interpret the findings, making the power curve analysis more practical.

While our initial aim was to achieve 80% power with five replicates, logistical constraints restricted us to four replicates, thereby marginally reducing our test power. This limitation has implications, potentially obscuring smaller treatment effects and inflating the maximum detectable difference between treatment means to an anticipated 50 calories burned.

Interestingly, post hoc analysis employing Tukey’s Honestly Significant Difference (HSD) test revealed a 95% confidence interval for mean differences, suggesting a maximum discrepancy of around 30 calories—a notable contrast to the initially projected 50 calories in our sample size calculation. This discrepancy hints at the possibility of our study having lower power than originally estimated.

However, given the complexities involved in our sample size calculation and the assumptions underlying it, there remains a possibility that our experiment did indeed achieve the targeted 80% power. To address any potential shortcomings, a viable solution would involve increasing the sample size in subsequent analyses or using computer simulation methods to obtain a more accurate sample size calculation.

Future experiments could benefit from employing factorial designs with a broader array of factors to explore the intricate interactions influencing caloric expenditure more comprehensively. Additionally, acknowledging that the blocking factors accounted for much of the variation and had large Mean Squares suggests that they could be explored further as independent variables in subsequent experiments. Considering the multitude of factors influencing caloric burn, implementing a 2^k factorial design may serve as an efficient screening tool to identify the most influential factors for further investigation, which then can be used as the basis for further experimentation. Furthermore, the significant differences observed among treatment levels underscore the potential effectiveness of tailored exercise regimens in maximizing caloric expenditure, with Treatment Level B showing promise in this regard.

Importantly, the absence of issues with residuals indicates that the model assumptions were met, enhancing confidence in the validity of the study's findings. Nevertheless, it's essential to acknowledge that this study represents a preliminary exploration into caloric expenditure and is not without its limitations. Future research and experimentation should aim to address these constraints and delve deeper into understanding the multifaceted nature of caloric expenditure during exercise. In summary, while this study offers valuable insights, its limitations highlight the need for further investigation into the complex relationship of factors influencing caloric expenditure, ultimately advancing our understanding of optimal exercise strategies for promoting energy expenditure and overall health.

6. Appendix

- Reach out to author for R code
- Data snapshot used in analysis

Duration	Intensity	TRT	Cal. Burned
30	High	A	316
30	Medium	B	342
10	Medium	C	113
10	Low	A	86
20	Low	B	188
20	High	C	194
30	Low	C	320
10	High	B	139
20	Medium	A	207
30	Low	C	280
20	Medium	A	172
30	High	A	284
30	Medium	B	319
10	High	B	120
20	Low	B	204
20	High	C	181

Table 1: Example of data used for experiment

- Summary Statistics of Cal. Burned by Treatment Level

```
data$TRT: A
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
   86.0    97.5   208.0   200.6   288.8   324.0
---------------------------------------------------
data$TRT: B
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
  120.0   138.5   195.0   223.2   324.8   355.0
---------------------------------------------------
data$TRT: C
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
   83.0   110.2   197.0   195.7   280.5   320.0
```

Figure 8: Summary Statistics of Cal. Burned by Treatment Level

- Boxplot of Cal. Burned by Treatment Level

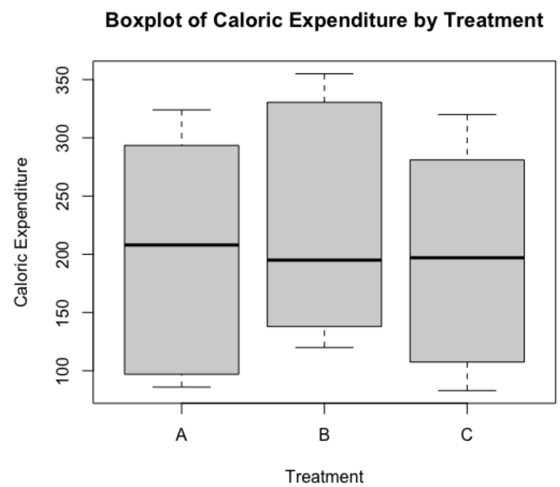

Figure 8: Boxplot of Cal. Burned by Treatment Level

- Residual vs. Fitted Plot for no interaction between treatment and block effects assumption for ANOVA model

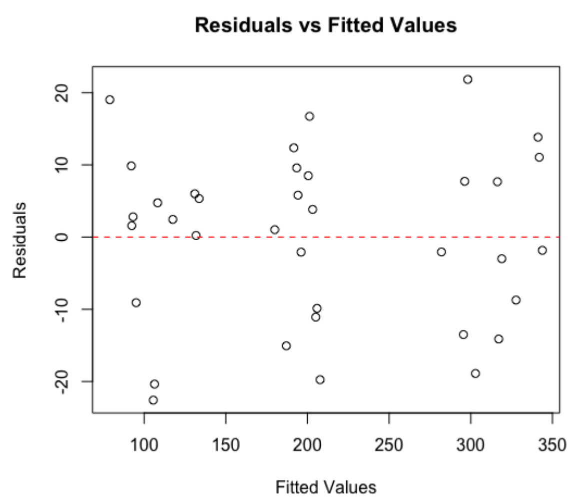

Figure 9: Residual vs. Fitted Value Plot showing adherence to no interaction assumption