Impact of Sleep and Training Intensity on Recovery and the Subsequent Day Strain

Jonas Freeman
Harvard University
Dept. of Statistics

Dries Rooryck
Harvard University
Dept. of Statistics

Eva Harris
Harvard University
Dept. of Statistics

Caden Woodall
Harvard University
Dept. of Statistics

Sam Lindman
Harvard University
Dept. of Statistics

1 Introduction & Motivation

Being healthy is more than a mindset; it is a lifestyle, and people commit to this lifestyle to varying degrees. From elite athletes to the "gym-bro," everyone who takes part in this lifestyle strives for optimal results. The athletic mindset will take you far, but you will only see an impact by training the body as well. However, it is impossible to see from the outside exactly what is happening within your body internally. In the present day, even 1% gains are celebrated achievements among professional and Olympic-level athletes. But if you are already performing incredibly well, how do you become the best that your physiology will allow? This is where WHOOP comes in.

WHOOP is one of the best fitness tracking bands in the world. All of the data is collected continuously over time, and the app interface uses extensive visuals to display this data. Its founder, Will Ahmed, is a Harvard alum who used his time at school to research the body's response to cardiovascular load and physiological strain. He believed that there was a better way to approach training geared toward maximizing performance. The band syncs roughly 100 times per second, creating a continuous dataset that provides valuable insights to allow the user to optimize their training. [2]

In a 2016 study, the Activity Strain metric of WHOOP was proven to better predict the performance of MLB pitchers than counting the number of pitches they threw (the standard practice at the time for MLB). WHOOP's Activity Strain is a measure of the intensity and duration of physical exertion, expressed on a scale of 0 to 21. It primarily considers cardiovascular load, heart rate zones, and the duration of the activity to derive a composite score. In a piece written by Will Ahmed himself, titled "WHOOP Approved for In-Game Use in Major League Baseball," WHOOP advertises that following this study, it was the first continuous tracking device of its kind to be allowed for use in an MLB game. As an example of its application, the piece illustrated the following scenario: "Clubs could find out that Player X is no longer effective when his Day Strain hits a certain number, regardless of whether he's thrown 50 pitches or 100 pitches to that point." [1]

However, there is more to predicting performance than Activity Strain. For instance, while optimizing the load on your muscles is a critical way of looking at the game, what about the recovery leading into the game? Every day, muscular structures degrade at a cellular level, and every night, they are rebuilt. This process generally occurs during restorative sleep. Restorative sleep is the combination of deep sleep (SWS) and REM sleep. Deep sleep (SWS) is known for the physical restoration that occurs throughout the body overnight. REM sleep is known as the mentally restorative stage of sleep, during which short-term memories are converted into long-term memories. [3]

In this project, we will address these findings by attempting to use other athletic parameters to predict the amount of time spent in deep sleep relative to overall sleep. We hypothesize that there is a relationship between this ratio and the load that workouts (or strain) place on the body.

2 Data & EDA

2.1 EDA & Hypothesis

The goal is to predict how much of the total sleep is spent in deep sleep (SWS) based on workout metrics, heart rate, and nap data. In our exploratory analysis, we decided to test whether we needed **WHOOP's Activity Strain** ("Activity Strain") in our model. The algorithm WHOOP uses is proprietary, but we needed to determine whether other data could be used to strongly predict this measure. If this is possible, then these data would be collinear. In the case that Activity Strain is collinear with another metric in the dataset, we could remove Activity Strain to avoid redundancy.

To predict Activity Strain, we created a linear regression regression that had Activity Strain as our a dependent variable. Because our goal was to test the redundancy of including this metric, we used all variables that were not distance-related or time-related as our exploratory variables. This meant excluding GPS.enabled, Distance..meters., Altitude.gain..meters., Altitude.change..meters., Cycle.start.time, Cycle.end.time, Cycle.timezone, Workout.start.time, Workout.end.time, and Activity.name.

Linear Regression Model for Activity Strain

$$\text{Activity Strain} \sim \left(\begin{array}{r} \text{Heart Rate Avg (bpm)} + \text{Heart Rate Max (bpm)} + \text{Duration (min)} + \\ \text{Total Energy Burned (cal)} + \text{Weighted Avg HR} + \text{HR Zone 0} + \text{HR Zone 1} + \\ \text{HR Zone 2} + \text{HR Zone 3} + \text{HR Zone 4} + \text{HR Zone 5} \end{array} \right) \tag{1}$$

The resulting model results of this model had an adjusted R^2 value of 0.95. Thus, **95%** of the variability in Activity Strain could be explained by the reduced dataset. This high level of explanatory power suggests that the simplified model captures the essential predictors, such as heart rate data (max heart rate, average heart rate, and heart rate zones) and workout duration. We then decided to check the observe the correlation matrix for of the workouts dataset for further insights about how our predictors relate to one another.

Figure 1 shows the correlation matrix for the workouts dataset.

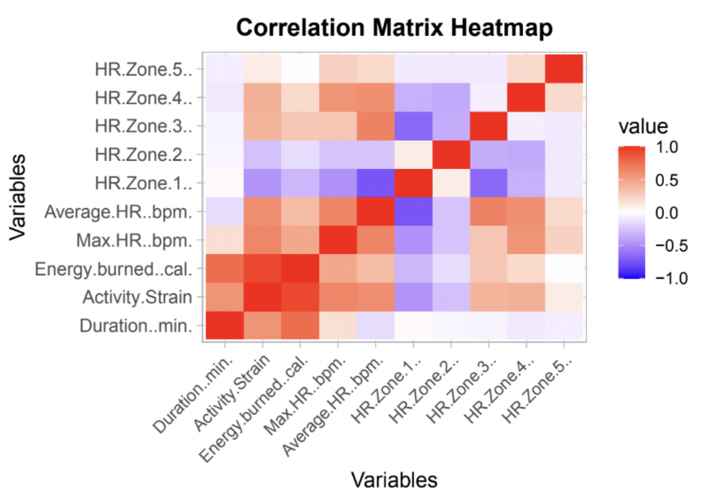

Figure 1: Correlation Matrix of the Workouts Dataset.

This correlation analysis of the workouts dataset was conducted using a threshold of 0.85 to identify highly correlated data. The resulting correlation graph revealed a strong correlation between Activity Strain and Energy Burned. To avoid issues of collinearity, the Activity Strain column was dropped, as Energy Burned serves as a reliable empirical measure of physical exertion.

To ensure compatibility with RStudio, we needed to clean the dataset and align the data types in the CSV files with the analysis environment.

Some common issues observed were tracking errors, such as incorrect activity detection and fragmented sleep data. Fortunately, most of these errors could be corrected manually through the app before exporting the data into CSV format. In the event of extraneous data, we cleaned it accordingly.

2.2 Data Collection & Cleaning

Originally, data was provided in three categories: **Workouts.csv** (covering details like workout duration, activity type, strain, energy burned, and heart rate zones), **Sleeps.csv** (including sleep cycle durations, efficiency, debt, and consistency), and **Physiological_Cycles.csv** (capturing recovery, resting heart rate, skin temperature, blood oxygen levels, and strain). While the data export process was simple, There was significant redundancy in the data, which required extensive cleaning. Because of this, we decided to create **one wrangled dataset that only included data from `workouts.csv` and `sleeps.csv`**. This final wrangled dataset contained 33 variables.

We chose to take a time-series approach. In the context of time series analysis, each day represents a single observation, allowing for the systematic tracking of workout metrics over time. We decided to create a dataframe that covered the days that Eva had a workout logged (and would be accounted for in workouts.csv), a dataframe that gathered the days that she did not have a workout logged, and then combine the two so that we have one, comprehensive dataset.

2.2.1 Preparing the Exploratory Variables

We first filtered days that included workouts based on workouts.csv. Columns with no data or redundant information were dropped, and the Activity.name variable was converted to a factor for categorical analysis. For clarity and interpretability, we set a floor value on the lowest number of observations of a workout. Activities labeled as "Other" (with only one observation) and "Activity" (with 65 observations) were removed. The Workout.start.time was converted to a Date type, and only the date portion was extracted to aggregate workout metrics by day.

Each of these daily workout summaries included several key metrics: `Total Duration`, which represents the sum of workout durations for each day, and `Total Energy Burned`, which is the total calories burned during workouts on that day. To account for workout duration when assessing heart rate, a `Weighted Average Heart Rate` was calculated, giving more weight to longer workouts. Additionally, weighted averages for each heart rate zone (1 through 5) were derived to capture the distribution of time spent in different heart rate intensities, again weighted by workout duration. By weighting the time spent in each HR zone by workout duration, you account for how long you maintained specific intensity levels across different workouts. This helps balance the influence of longer workouts on the daily average, which intuitively makes sense to impact the need for Deep SWS sleep. The `Max Heart Rate` was recorded as the highest heart rate for each day, reflecting peak exertion levels.

To gain a comprehensive view of workout types, the dataset includes counts for each activity, where each activity type (e.g., running, cycling) is tallied based on boolean indicators (0 or 1). A new metric, `Heart Rate Zone 0`, was introduced to quantify the percentage of time spent in a resting or low-intensity state. This was calculated by subtracting the sum of the percentages of heart rate zones 1 through 5 from 100, with the result rounded to two decimal places. These daily aggregates form a robust time series dataset, facilitating the analysis of workout trends, intensity distributions, and recovery patterns over time. This structured approach to summarizing daily workout data ensures that each observation captures a comprehensive snapshot of physical activity, allowing for meaningful insights into performance and recovery.

The other part to time series analysis is ensuring a complete and continuous date range. We took additional steps to ensure that this assumption was held. The process begins by generating a sequence of dates from the earliest workout date to the latest workout date in the `daily_workouts` dataset. This ensures that no dates are skipped, providing a consistent timeline.

We then used a new dataframe to represent rest days. For each date in the range, default values are assigned to indicate no physical activity and resting heart rate conditions. Metrics like `Total Duration` and `Total Energy Burned` are set to 0, reflecting no workout activity. Heart rate-related variables, such as

Weighted Average Heart Rate and Max Heart Rate, are assigned a resting value of 57 bpm. Additionally, the weighted average heart rate zones (Zones 1 through 5) are set to 0%, while Heart Rate Zone 0 is set to 100%, indicating that all heart rate activity is in the resting range. The activity count variables (e.g., Total Running, Total Cycling) are also set to 0, indicating no activity was performed on these days. By incorporating these rest days into the dataset, the continuity of the time series is maintained, which is crucial for generating accurate insights and visualizations.

By combining the dataframe of rest days with the original data from workouts.csv, we were able to make sure that the days with recorded workouts retain their actual metrics, such as higher maximum heart rates and real weighted heart rate zones. Our use of default placeholders in the dataframe of rest days did not overwrite or exclude higher heart rate values in the original dataset, but instead allowed for a clear distinction between workout days and rest days. After merging, the dataset was processed to remove duplicate entries, which kept only the first occurrence of a workout that was logged. To maintain the integrity of the time series, the dataset was then sorted chronologically.

This final wrangled dataset contains a comprehensive set of variables that provide a detailed view of daily workout and heart rate metrics. These variables include `Workout.start.date` to represent each observation, `Total.Duration.min` for the total workout duration in minutes, and `Total.Energy.burned.cal` for the total calories burned during workouts. The dataset also includes `Weighted.Avg.HR`, which reflects the weighted average heart rate based on the duration of each workout, and detailed heart rate zone metrics such as `Weighted.Avg.HR.Zone.0` through `Weighted.Avg.HR.Zone.5`, indicating the percentage of time spent in different heart rate intensity zones. Additionally, `Max.HR.bpm` captures the highest heart rate recorded for each day, reflecting peak exertion levels. The dataset also tracks the frequency of various activities with boolean counts for each type, including `Total.Assault_Bike`, `Total.Cycling`, `Total.Elliptical`, `Total.Functional_Fitness`, `Total.Hiking_Rucking`, `Total.Powerlifting`, `Total.Rowing`, `Total.Running`, `Total.Spin`, `Total.Stairmaster`, `Total.Walking`, `Total.Weightlifting`, and `Total.Yoga`. These activity counts help in understanding the distribution and variety of workouts performed over time.

2.2.2 Preparing the Response Variable

We removed `Cycle.start.time`, `Cycle.end.time`, `Cycle.timezone` because these are time-related metrics that are not in line with our goal. Our goal is to predict the ratio of time spent in Deep (SWS) Sleep to the overall amount of time spent asleep. We plotted a correlation matric of sleep.csv to understand how the variables relate to one another better and to justify other changes to the dataset. In this, we found that there were four highly correlated variables.

Figure 2 shows the correlation matrix for the sleep dataset.

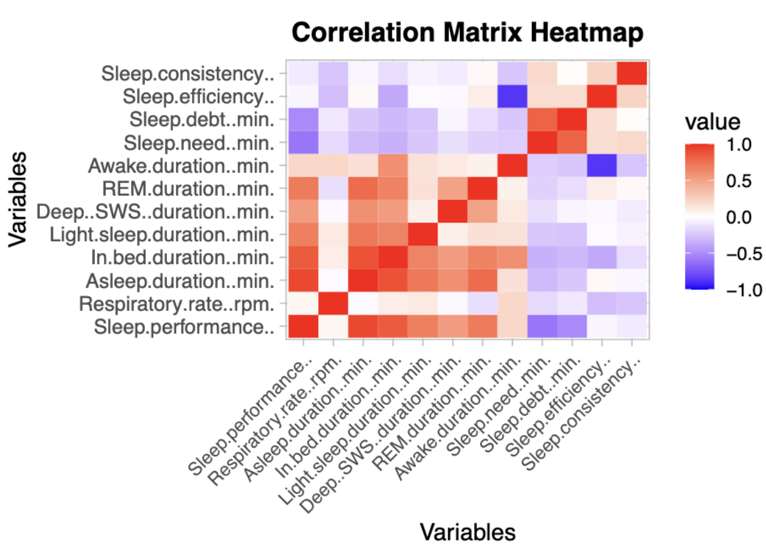

Figure 2: Correlation Matrix of the Sleep Dataset.

This correlation analysis of the sleep dataset was conducted using a threshold of 0.85 to identify highly correlated data. The resulting correlation graph revealed a strong correlation between four variable pairs: asleep duration and sleep performance, in-bed duration and sleep performance, in-bed duration and asleep duration, and sleep efficiency and awake duration. To avoid issues of collinearity, we dropped in-bed duration and sleep performance because asleep duration proved to be a reliable empirical measure. Additionally, while not caught by the threshold, sleep debt and sleep need appear to be highly corrrelated in the graph above. We decided to drop sleep debt as a result.

In-bed duration refers to the total time spent in bed, whether asleep or awake. Sleep performance is a metric calculated by a proprietary WHOOP algorithm that takes into account how well you were able to meet your body's need for sleep, incorporating sleep debt and naps. Being able to reproduce these values simply with asleep duration simplifies things significantly. Similarly, we also dropped sleep efficiency because awake duration proved to be a reliable empirical measure. Sleep efficiency is a metric calculated by a proprietary WHOOP algorithm to measure how fast you go to sleep.

Nap statistics were then calculated by converting `Wake.onset` to `Wake.date` (date-only format) and grouping by `Wake.date`. For each date, `Nap.count` was computed as the number of distinct wake-up times minus one, and `Nap.duration` was the total sleep duration minus the longest sleep (main sleep). If no naps occurred, `Nap.duration` was set to zero. This provided clear daily measures of nap frequency and duration.

The main sleep session for each day was identified as the entry with the maximum `Asleep.duration..min.`, retaining the first occurrence in case of ties. To align sleep with workout data, a `Sleep.date` column was created by subtracting one day from `Wake.onset`, reflecting the night the user went to sleep. The main sleep data was merged with nap data using a left join on `Sleep.date` and `Wake.date`, filtering out rows with missing nap statistics. The final dataset was grouped by the wake-up date to organize sleep data by the day the user woke up. The resulting dataset includes primary sleep duration, nap count, and total nap duration, all aligned for consistency with workout data.

2.2.3 Combining Exploratory and Response Data

Next, we used the date of each observation to perform an inner join that combines the dataset that contained our exploratory variables with the dataset that contained our response variables. Additionally, to streamline this dataset, time-related columns such as `Sleep.onset`, `Wake.onset`, and `as.Date(Wake.onset)` were removed, as well as the count for each type of workout. The number of times Eva logged each type of workout is extraneous because it is the heart rate zones that trains the cardiovascular system, irrespective of the actual exercise or method used to achieve this heart rate.

The final dataset had the following columns: respiratory rate, asleep duration, light sleep duration, deep sleep (SWS) duration, REM duration, awake duration, sleep need, sleep debt, sleep consistency, nap count, nap duration, total workout duration, total energy burned, weighted average heart rate, weighted average heart rate zones 0 through 5, and maximum heart rate.

and can be represented by this correlation matrix:

Figure 3 shows the correlation matrix for the cleaned dataset.

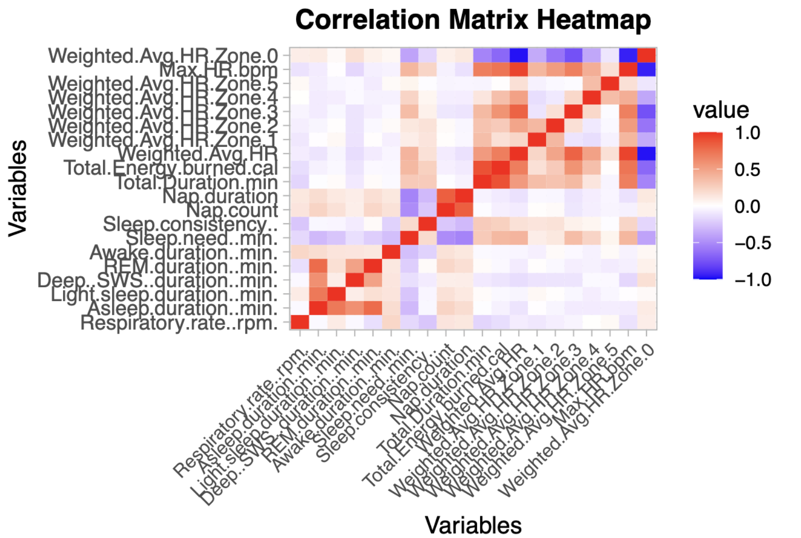

Figure 3: Correlation Matrix of the Cleaned Dataset.

3 Methods

3.1 Methods for Inference Models

3.1.1 Assumptions for Baseline Model

Before we began to construct our the baseline linear regression model, we first checked the standard assumptions: Existence, Linearity, Independence, and Homoscedasticity (ELIH). The results of these checks are visualized through residual diagnostics, including Residuals vs. Fitted plots, Q-Q plots of residuals, and histograms.

- **Linearity and Homoscedasticity:** The Residuals vs. Fitted plot *(Figure 4)* shows that the residuals are largely centered around zero with no clear non-linear pattern. However, there is slight heteroscedasticity, as evidenced by uneven spread near the center of the fitted values.
- **Normality of Residuals:** The Q-Q plot *(Figure 4)* indicates that the residuals follow an approximately normal distribution, with minor deviations in the tails caused by a few outliers.
- **Independence of Residuals:** The Residuals vs. Total Duration plot *(Figure 5)* demonstrates that residuals are randomly scattered, suggesting no relationship between the residuals and predictor values.
- **Residual Distribution:** The histogram of residuals *(Figure 5)* shows a roughly bell-shaped curve, supporting the assumption of normally distributed errors.

The residual diagnostics confirm that the assumptions of linearity, independence, and normality are approximately satisfied. Minor deviations in homoscedasticity and the presence of a few outliers should be addressed in future models through transformations or robust regression techniques.

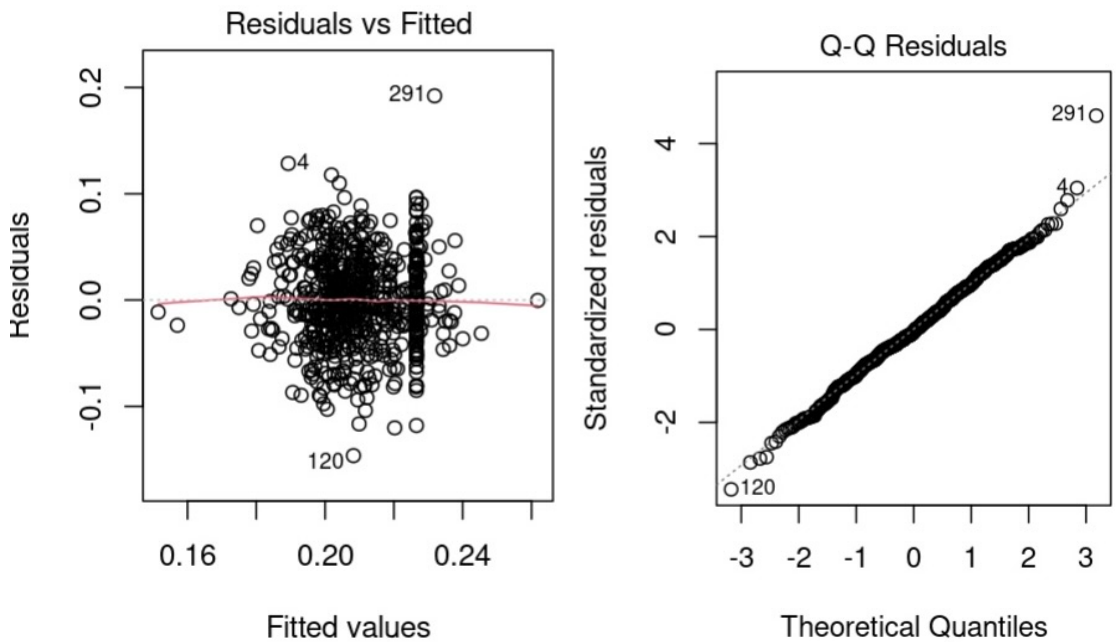

Figure 4: Residuals vs. Fitted (left) and Q-Q Plot of Residuals (right).

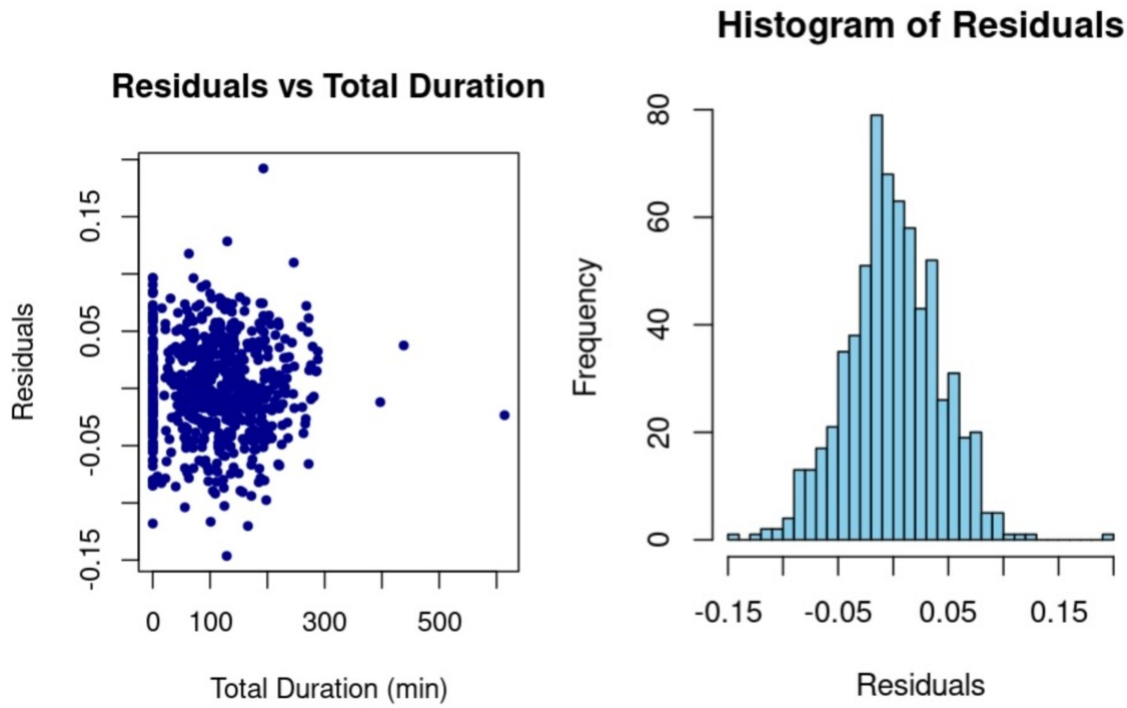

Figure 5: Residuals vs. Total Duration (left) and Histogram of Residuals (right).

3.1.2 More Sophisticated Models

We implemented one baseline model and two sophisticated models for inference in our analysis. The baseline model was a generalized linear regression designed to predict the ratio of SWS (Deep) sleep to overall sleep. The **adjusted** R^2 **value** of the baseline model was 0.347**.** This prediction relied on a comprehensive set of variables that included respiratory rate, total sleep duration, deep sleep duration, nap count, nap duration, total workout duration, total energy burned, and heart rate data including the weighted average heart rate (`Weighted.Avg.HR`), wieghted heart rate zone distributions (`Weighted.Avg.HR.Zone.0` through `Weighted.Avg.HR.Zone.5`) and maximum heart rate (`Max.HR.bpm`). There were no interaction terms. The aim for this diversity was to provide a comprehensive basis for predicting the ratio of deep sleep.

Our first attempt at a more sophisticated model included an interaction term with `Total.Energy.burned.cal`. The model `model_rr` is a linear regression with interaction effects that is designed to predict the ratio of time spent in deep sleep (SWS) to the total amount of time spent asleep. The **adjusted** R^2 value of the interaction effects this model was 0.351, which is significantly higher than our baseline model (0.347) It uses the total energy burned during workouts and interactions with all other predictors in the dataset. The response variable, Deep..SWS..duration..min. / Asleep.duration..min., represents the proportion of sleep spent in deep sleep, which is essential for physical recovery. The predictor Total.Energy.burned.cal captures the total calories burned during the day, reflecting workout intensity or duration.

By including the interactive term, the model incorporates both the direct effect of total energy burned on the deep sleep ratio and the interactions between energy burned and every other variable in the dataset. These interactions allow the model to assess how the relationship between each predictor and the deep sleep ratio changes depending on the amount of energy burned. For example, the effect that average heart rate or nap duration may have on deep sleep may vary, based on how much energy was expended that day.

With our second, more sophisticated model, we developed a linear regression that predicts the deep sleep ratio by incorporating all main effects and two-way interactions between the predictors. While this approach captures more intricate relationships, it also introduces risks such as overfitting and reduced interpretability due to the interaction terms. The **adjusted** R^2 **value** for the full effects model is 0.364. It makes sense that the most complex model would be a suitable candidate for backward stepwise selection. Backward step selection systematically simplifies the model by eliminating the insignificant interaction terms. To evaluate the quality of the models at each step of this process, we employed the Akaike Information Criterion (AIC), which assesses model performance relative to other potential models.

Originally, we had a higher R^2 value in the interaction model. This did not make sense, so we reexamined our code and made corrections accordingly. To ensure that we had most optimal model through step selection, we conducted the stepwise selection process twice: first using the interactive effects model and then using the full effects model. The interactive effects model yielded an AIC of **-4415.95**, which is notably lower than the AIC of **-4295.39** obtained for the Full Effects Model. This can be used to conclude that the interactive effects model yielded a superior balance of model fit and complexity. In both processes, backward selection produced a higher R^2 value, so we have included the structure of both backward step selections below:

Backward, Interactive Effects Model: $\frac{\textbf{Deep SWS Duration (min)}}{\textbf{Asleep Duration (min)}}$

$$\sim \text{Total Energy Burned (cal)} \times$$

$$\left(\begin{array}{r}\text{Respiratory Rate (rpm)} + \text{Light Sleep Duration (min)} + \text{REM Duration (min)} + \text{Awake Duration (min)}+\\ \text{Sleep Need (min)} + \text{Sleep Debt (min)} + \text{Sleep Consistency} + \text{Nap Count} + \text{Nap Duration}+\\ \text{Total Duration (min)} + \text{Total Energy Burned (cal)} + \text{Weighted Avg HR} + \text{HR Zone 1}+\\ \text{HR Zone 2} + \text{HR Zone 3} + \text{HR Zone 4} + \text{HR Zone 5} + \text{Max HR (bpm)} + \text{HR Zone 0}\end{array}\right) \tag{2}$$

$$\left(\begin{array}{r}\text{Respiratory Rate (rpm)} + \text{Light Sleep Duration (min)} + \text{REM Duration (min)} + \text{Awake Duration (min)}+\\ \text{Sleep Need (min)} + \text{Sleep Debt (min)} + \text{Sleep Consistency} + \text{Nap Count} + \text{Nap Duration}+\\ \text{Total Duration (min)} + \text{Total Energy Burned (cal)} + \text{Weighted Avg HR} + \text{HR Zone 1}+\\ \text{HR Zone 2} + \text{HR Zone 3} + \text{HR Zone 4} + \text{HR Zone 5} + \text{Max HR (bpm)} + \text{HR Zone 0}\end{array}\right) \tag{3}$$

Backward, Full Effects Model: $\frac{\textbf{Deep SWS Duration (min)}}{\textbf{Asleep Duration (min)}}$

$$\sim \left(\begin{array}{r}\text{Respiratory Rate (rpm)} + \text{Light Sleep Duration (min)} + \text{REM Duration (min)} + \text{Awake Duration (min)}+\\ \text{Sleep Need (min)} + \text{Sleep Debt (min)} + \text{Sleep Consistency}+\\ \text{Nap Count} + \text{Nap Duration} + \text{Total Duration (min)}+\\ \text{Total Energy Burned (cal)} + \text{Weighted Avg HR} + \text{HR Zone 1}+\\ \text{HR Zone 2} + \text{HR Zone 3} + \text{HR Zone 4} + \text{HR Zone 5}+\\ \text{Max HR (bpm)} + \text{HR Zone 0}\end{array}\right)^2 \tag{4}$$

Table 1: Adjusted R^2 Values for Different Models

Model	**Adjusted** R^2
Baseline Model	0.347
Interaction Effects Model	0.351
Full Effects Model	0.364
Backward Stepwise (Interactive Effects Model)	0.362
Forward Stepwise (Interactive Effects Model)	0.353

The **adjusted** R^2 for the most optimal model (backward step selection of the interactive effects model), we observed a value of 0.362. The **adjusted** R^2 for the interactive effects model obtained through forward stepwise selection, was a value of 0.353. Overall, there was improvement over time as we refined the model. However, One potential hazard to be wary of is that interaction terms may exacerbate multicollinearity issues, especially if the predictors involved in the interaction are correlated. It is important to note that high multicollinearity can lead to inflated standard errors and less accurate coefficient estimates.

For formatting, the fitted/residuals plots and QQ plots for each of the Models referenced in this section to the ***Appendix***

3.2 Methods for Prediction Models

3.2.1 Data Splitting & Model Evaluation

3.2.2 Baseline Predictive Logistic Model

3.2.3 Constructing Other Models

4 Results

4.1 Inference

4.2 Prediction

4.3 Potential Limitations

5 Conclusion & Discussion

References

[1] Will Ahmed. Whoop approved for in-game use in major league baseball. *WHOOP*, March 2017.

[2] CPerlman. Whoop: Using data to target the elite athlete. *Digital Innovation and Transformation*, 2024. Accessed 17 Dec. 2024.

[3] WHOOP. What is restorative sleep? *WHOOP*, October 2023.

Appendix: Parameter Distributions & Model Plots

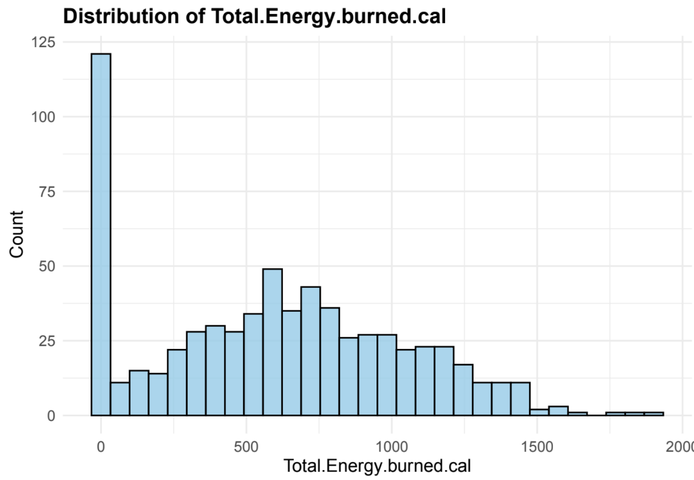

Figure 6: Distribution of Total.Energy.burned.cal

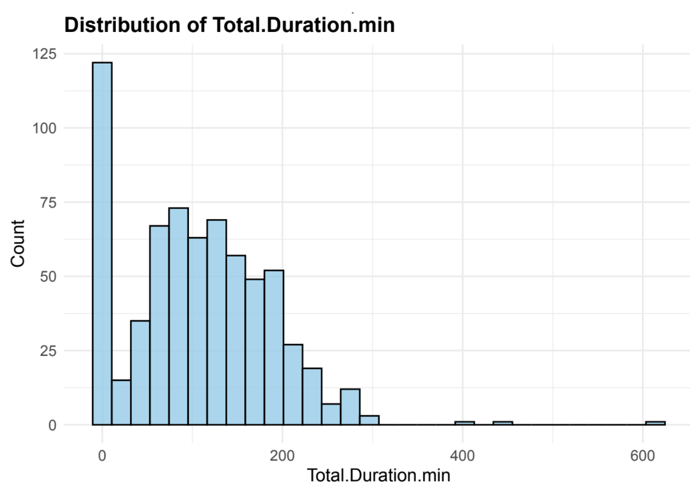

Figure 7: Distribution of Total.Duration.min

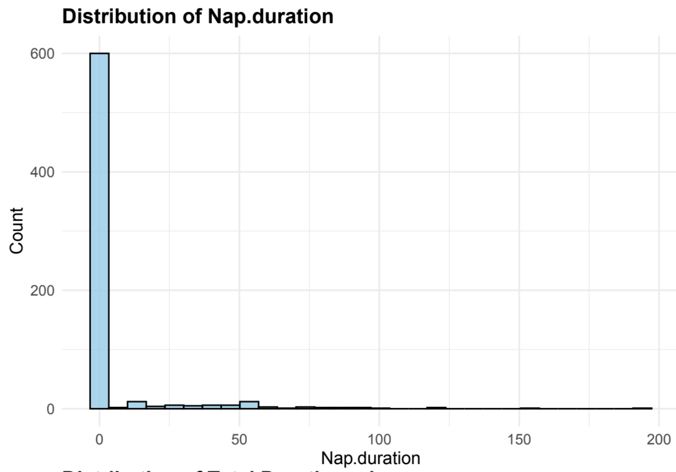

Figure 8: Distribution of Nap.duration

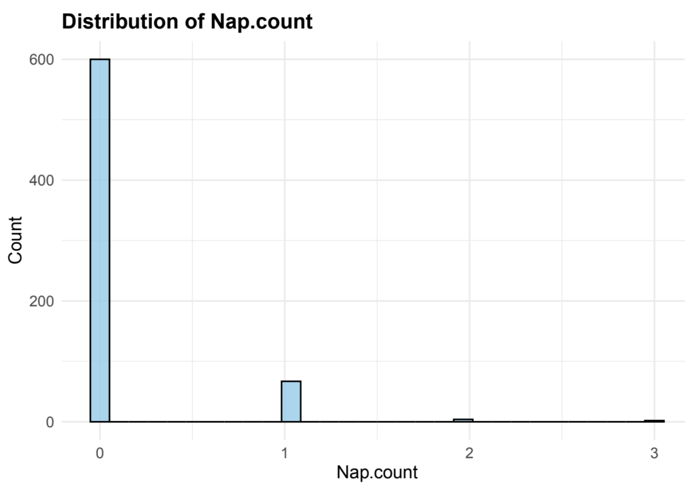

Figure 9: Distribution of Nap.count

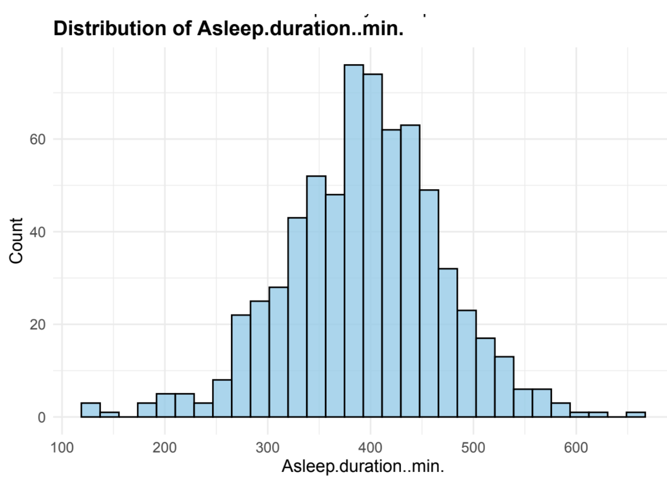

Figure 10: Distribution of Asleep.duration..min.

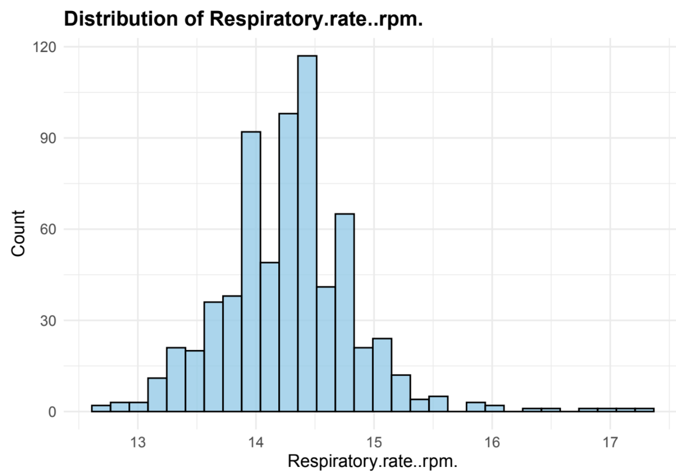

Figure 11: Distribution of Respiratory.rate..rpm.

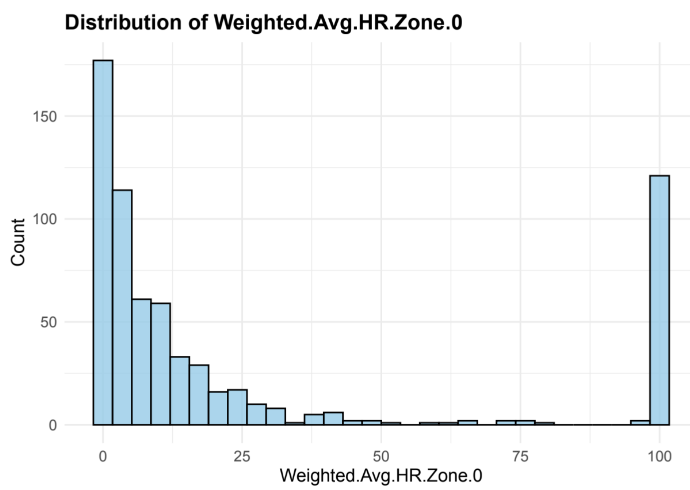

Figure 12: Distribution of Weighted.Avg.HR.Zone.0

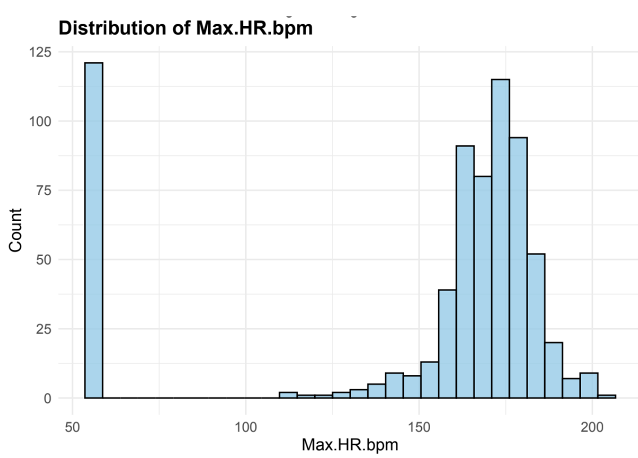

Figure 13: Distribution of Max.HR.bpm

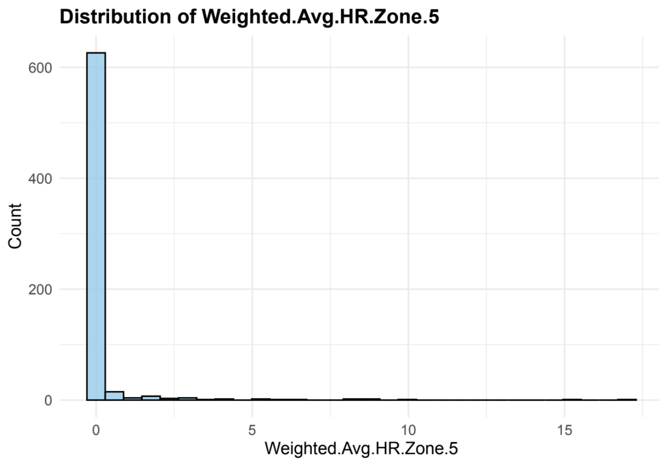

Figure 14: Distribution of Weighted.Avg.HR.Zone.5

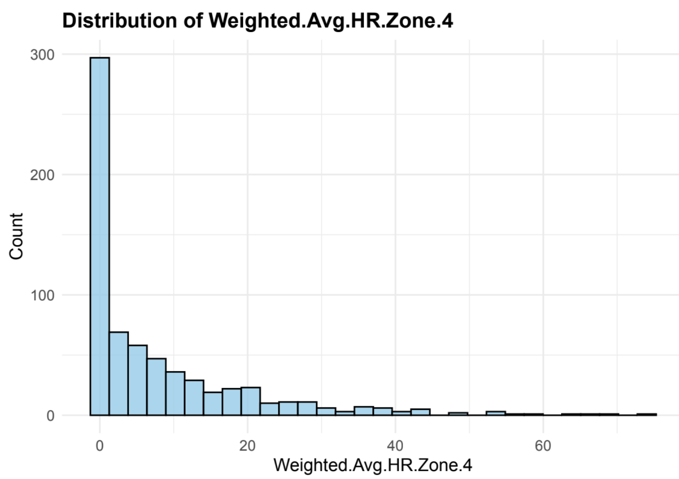

Figure 15: Distribution of Weighted.Avg.HR.Zone.4

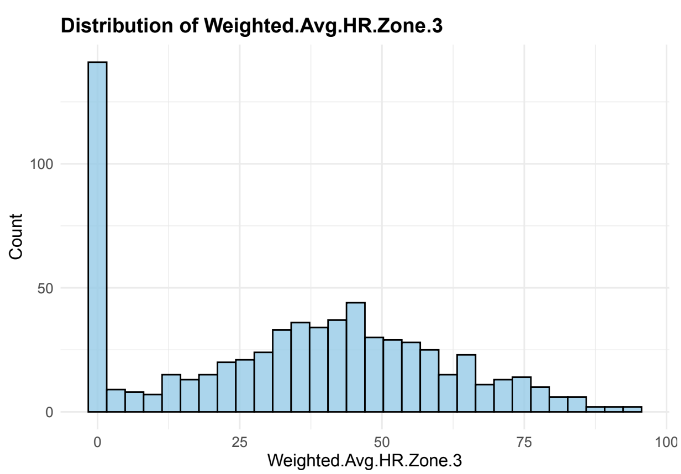

Figure 16: Distribution of Weighted.Avg.HR.Zone.3

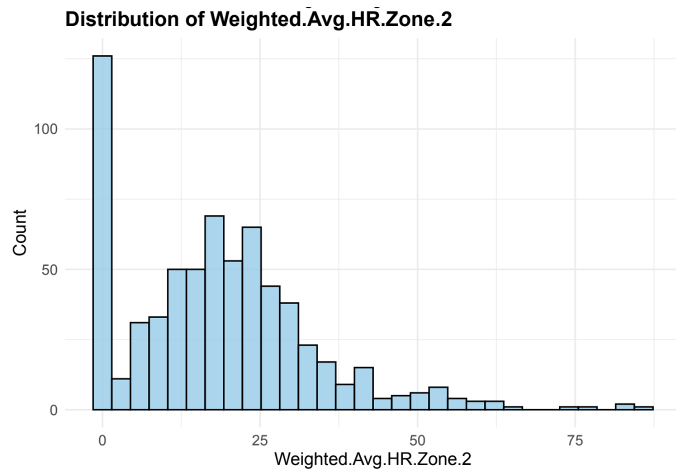

Figure 17: Distribution of Weighted.Avg.HR.Zone.2

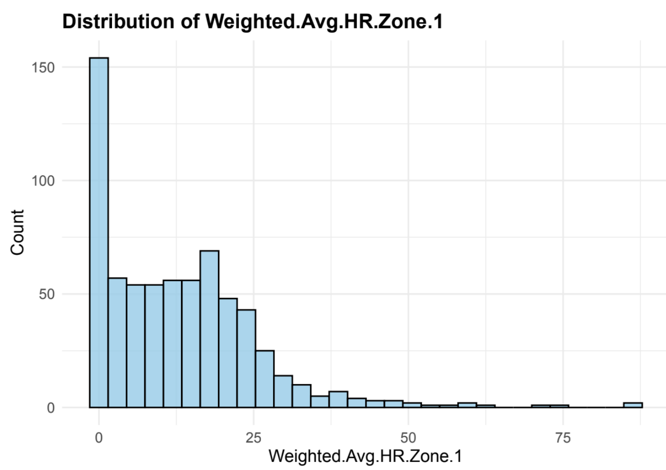

Figure 18: Distribution of Weighted.Avg.HR.Zone.1

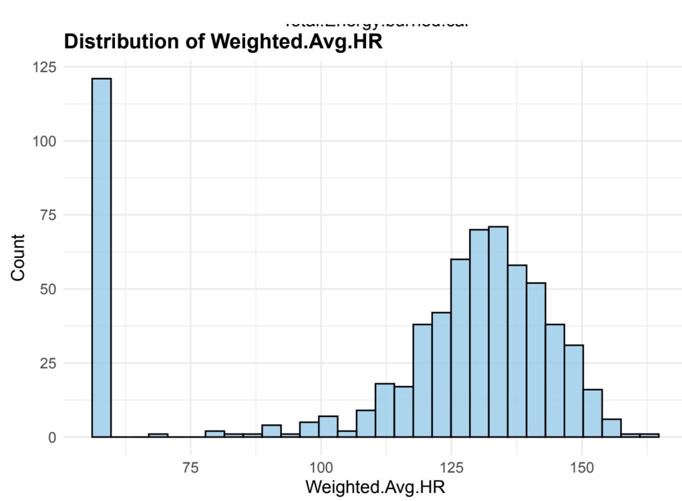

Figure 19: Distribution of Weighted.Avg.HR

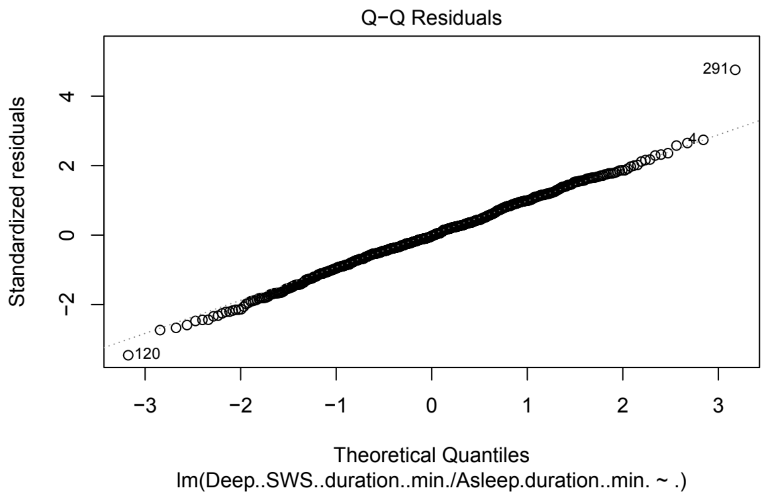

Figure 20: General Model: Residuals vs Fitted

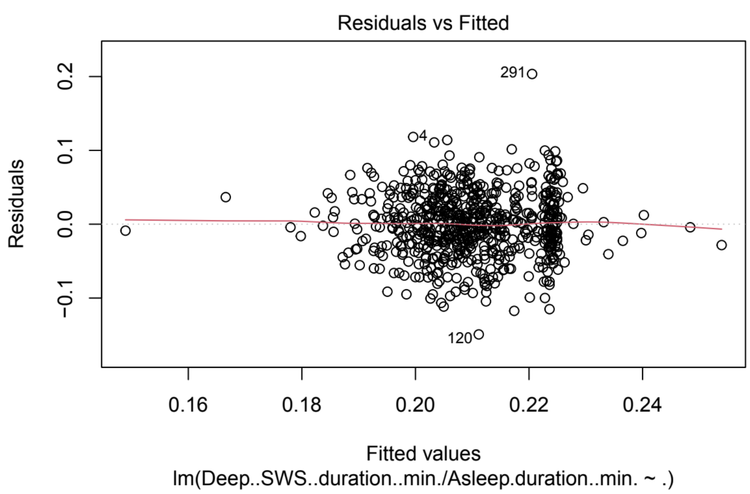

Figure 21: General Model: Q-Q Residuals

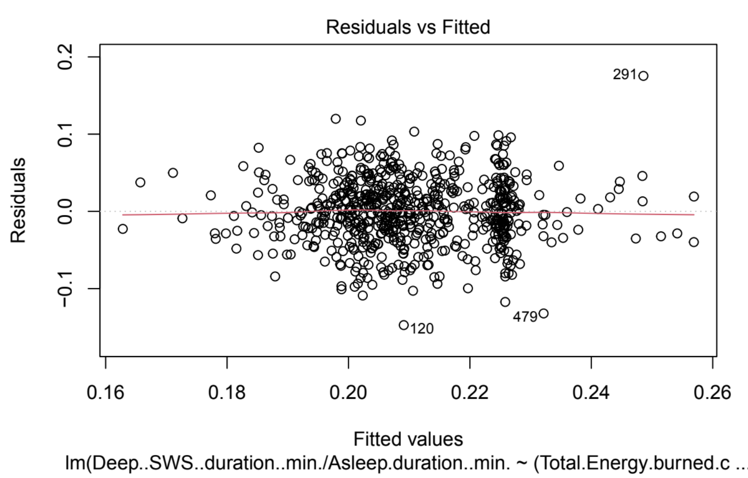

Figure 22: Interactions Model: Residuals vs Fitted

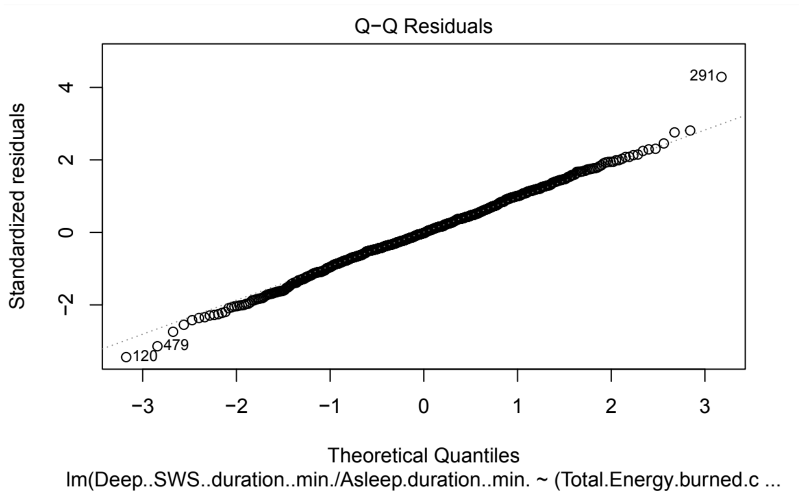

Figure 23: Interactions Model: Q-Q Residuals

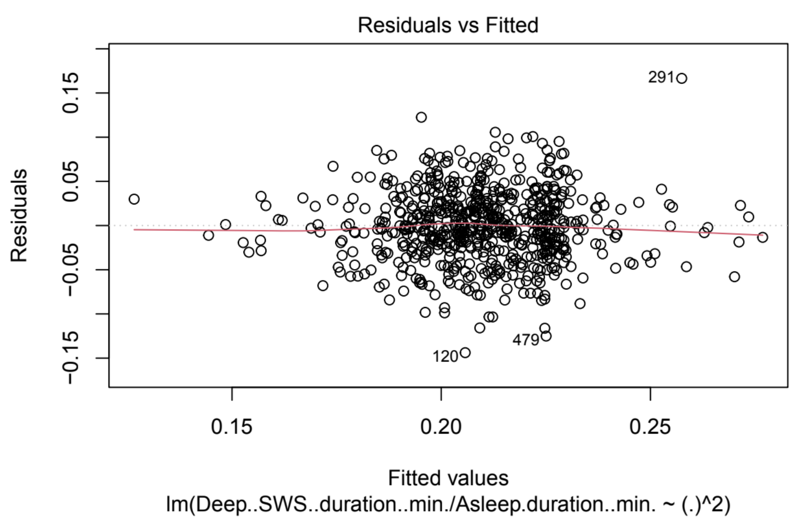

Figure 24: Full Model: Residuals vs Fitted

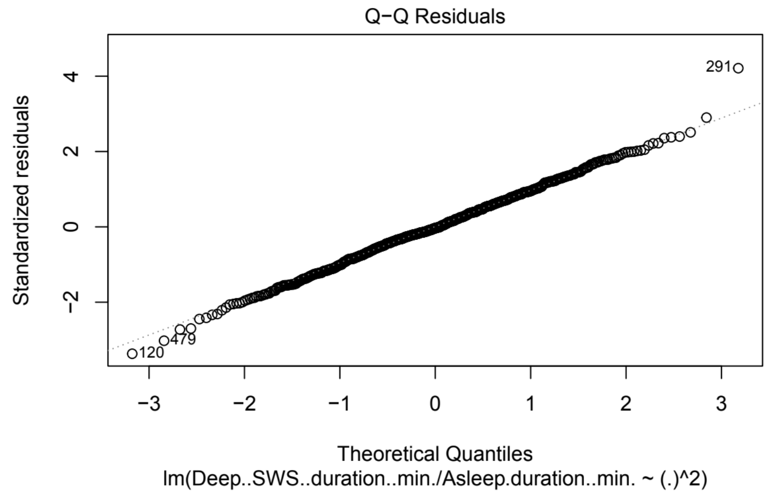

Figure 25: Full Model: Q-Q Residuals

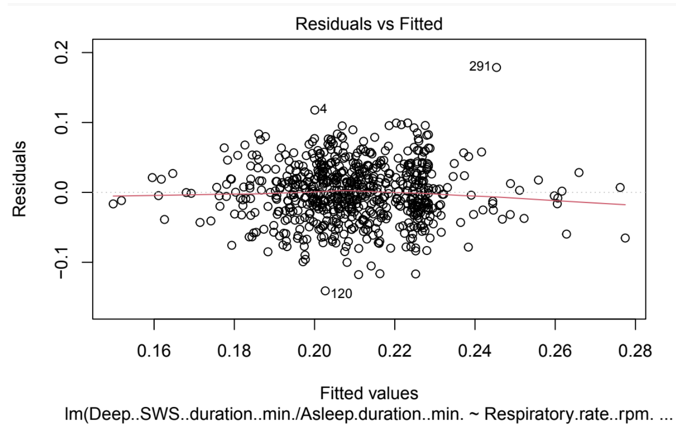

Figure 26: Model Selection (Backward): Residuals vs Fitted

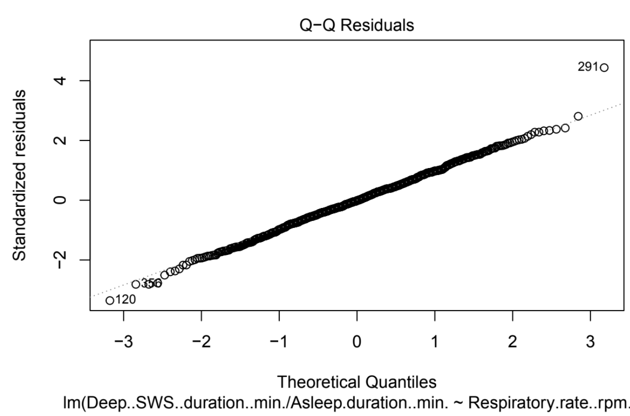

Figure 27: Model Selection (Backward): Q-Q Residuals

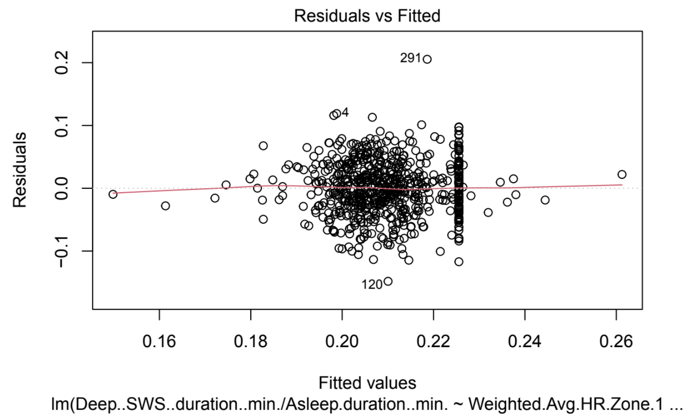

Figure 28: Model Selection (Forward): Residuals vs Fitted

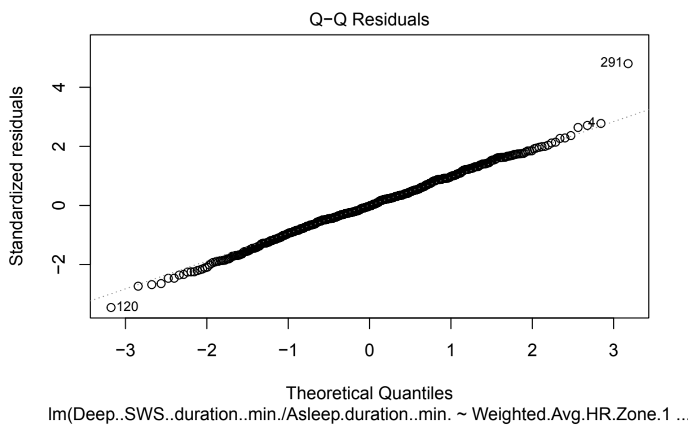

Figure 29: Model Selection (Forward): Q-Q Residuals

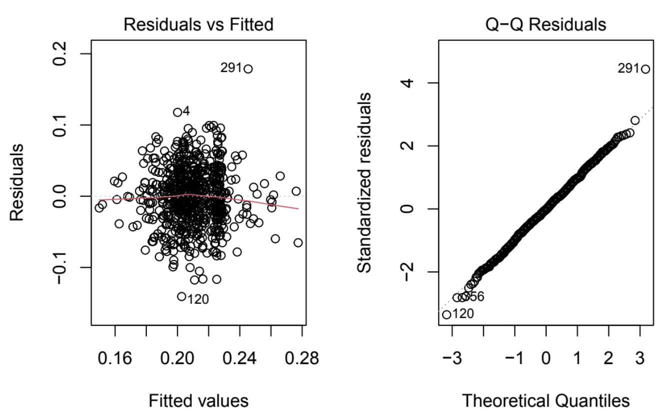

Figure 30: Prediction: Residuals vs Total Duration and Histogram of Residuals

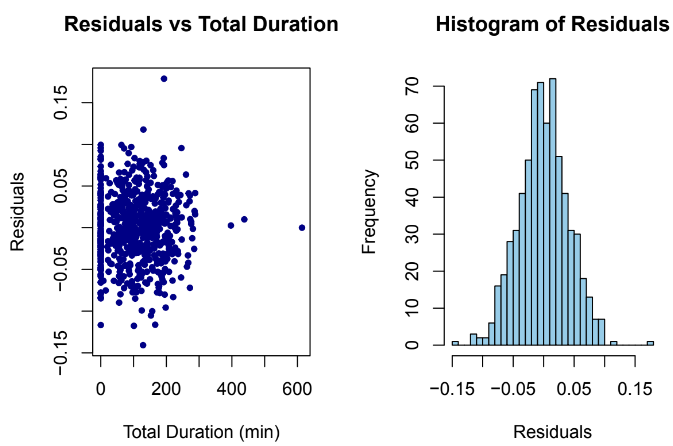

Figure 31: Prediction: Residuals vs Fitted Values and Q-Q Plot of Residuals