Springboard Intro to DS: Technical Capstone Paper

Joe Earnshaw

For a little over a year, I've been utilizing a wearable known as WHOOP. Names like Fitbit, Apple Watch and Garmin Vivosport are usually some of first names that come to mind when it comes to the wearables. While I believe that these technologies can promote the formation of healthier habits, WHOOP takes it to the next level.

This device is minimalist in nature (there is no screen interface directly on the device), the data and insights it provides pack quite a punch. Essentially what WHOOP does is it continuously tracks a set of health markers (like heart rate and heart-rate variability, which I will get into later) and provides this information via three primary metrics: Strain and Sleep Performance and Recovery. With this information, you are better able to understand the effects of your training while also adjusting it for days that your body may not be fully recovered to take on a substantial cardiovascular load (i.e. active recovery and rest days).

WHOOP Strain is a measure of an individual's total cardiovascular load over a specific time domain, which is then relayed to the user on a scale from zero to 21, with 0 being absolutely no activity and 21.0 being the maximal load one could attain in a day. It does not measure steps or the miles you ran -- instead by tracking heart rate during workouts it is able to gauge your physiological response to the stimulus and determine the impact that activity had on your body. Also, because you wear it continuously, it is able to track your response to non-training events like grocery shopping, cooking dinner, etc.

WHOOP Sleep Performance is the total amount of sleep you got on any given night divided by the amount of sleep you needed based off the strain you accumulated throughout that particular day. It is reported as a percentage from 0-100, with the goal being that you want to be as close to 100% as possible. Additional sleep-related markers the strap tracks include time spent in bed, time spent asleep, in the various stages of sleep (light, REM, and Deep), and the total number of sleep cycles.

WHOOP Recovery works the other side of the training equation i.e. it measures the body's ability to adapt to a training stimulus. This is primarily through tracking sleep in addition to heart-related measures like resting heart rate and heart rate variability (HRV). There is a growing body of evidence that HRV is a great, non-invasive measure of the body’s current state to accumulate cardiovascular load; by being able to track it, you have the potential to get better insights into your fitness/fatigue balance.

While WHOOP has provided me with amazing and actionable information, I wanted to go deeper. I've created a spreadsheet with daily observations (from the past 4 months) on the following variables:

* Strain score (on a scale from 0-21)
* Recovery score (as a %)
* Sleep Performance score (sleep your body needs vs. the sleep achieved each night, as a %)
* Max Heart Rate (highest heart rate for that specific day)
* Average Heart Rate (average heart rate from all activities for that day)
* Calories (approximate calories burned that day)
* Heart-Rate Variability (as measured during slow wave sleep and uses the metric RMSSD (Root Mean Square of Successive Differences)
* Resting Heart Rate (measured at complete rest during your deepest sleep each night)
* Time In Bed (in hrs.)
* Time Spent in Light Sleep stage (in hrs.)
* Time Spent in REM Sleep (in hrs.)
* Time Spent in Deep Sleep (in hrs.)
* Total Sleep (in hrs.)
* Total Sleep Cycles

My Springboard Mentor Jarus Singh and I started wondering: what if, with the information gathered from WHOOP over the past few months, we could predict variables such as Strain, Sleep Performance and Recovery to create a long-term workout plan, say for a week?

The benefits of this would be significant; you could plan exactly which days during the week you wanted to go all out, and the days where you take it a little easier in order to maximize overall performance. Additionally, it can help in terms of scheduling training around life events, after all you don't want to cut short, or even eliminate, a training session on a day where you're in the green zone and leaving potential gains on the table! Instead, if you know that you're going to have a work party on Thursday you can plan your training around that so that day is an active recovery/rest day and you won't have to worry about missing out on training while fully enjoying the party!

**Initial Exploratory Analysis**

My initial exploratory analysis showed some interesting trends, which ended up helping significantly in the model creation stage of this project. I began by getting a simple summary of the data so far, which included the min, 1st quartile, median, mean, 3rd quartile and max for each variable listed above.

My first observation was from trends in my total sleep. With a median of 7.190 hrs. and mean of 6.930 hrs., I would rate this 4-month period’s overall sleep performance as 'needs improvement'. Matthew Walker, a neuroscientist and director of the Center for Human Sleep Science at UC Berkeley, has stated that the vast majority of people need between seven and nine hours of sleep a night (at minimum). While I am technically within that range, I would still say the amount of sleep I am getting is inadequate, due to a training schedule that is anywhere from 1-3 hrs./ day, 5-6 times a week. Higher strain puts greater stress on the body and essentially the only time our bodies recover is when we sleep.

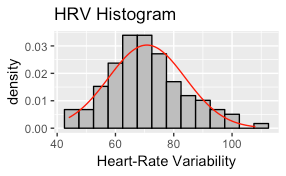
Then if you take a look at the 'recovery' variable you will see that this lack of sleep may indeed be having a negative impact on my performance. With a mean and median recovery score of 0.5902 and 0.6000 (expressed here as a decimal; in WHOOP it is displayed as a % out of 100, with 0% being the lowest score and 100% being the highest), respectively, it appears that my recovery's have not been all that great. In fact, this puts me in the 'yellow zone', where I have to be very conscious of not putting too much strain on my body that particular day. By seeing these numbers, I'm able to get actionable insights right away and the first key takeaway is that I need to create more time in my schedule for sleep, whether that is more time in bed at night or by utilizing naps during the day.

The next variable I want to look at is heart-rate variability. First, some background is necessary before taking a deeper dive. On a high level, heart-rate variability, or HRV, is a measure of irregularity in the heart rate ([1](https://www.whoop.com/the-locker/an-athletes-guide-to-hrv/?gclid=Cj0KCQjw3KzdBRDWARIsAIJ8TMRAgo8EOPlnujmFKGPoqavNn75jrJfOYJOKkG4aV5R9IpZmPOcDkwwaAnxiEALw_wcB)). Essentially, the heart doesn't have a perfect rhythm, nor does it beat in perfect intervals. Instead there is some variability in the time between heartbeats, which is called the RR-interval, and HRV is a function of the difference in the lengths of successive beats in a series of these intervals ([1](https://www.whoop.com/the-locker/an-athletes-guide-to-hrv/?gclid=Cj0KCQjw3KzdBRDWARIsAIJ8TMRAgo8EOPlnujmFKGPoqavNn75jrJfOYJOKkG4aV5R9IpZmPOcDkwwaAnxiEALw_wcB)).

Yet, what controls this variation? HRV is largely controlled by the autonomic nervous system (ANS), which then can be subdivided into two components, the sympathetic (i.e. fight-or-flight mechanism) and parasympathetic (i.e. relaxation or 'rest-and-digest' mechanism) nervous system ([2](https://www.health.harvard.edu/blog/heart-rate-variability-new-way-track-well-2017112212789)). The hypothalamus, a part of our brain that is constantly processing information, sends signals through the ANS to either stimulate or relax it. Now this is where HRV comes back into play as it is able to detect if there any imbalances between your fight-or-flight and rest-and-digest responses. Here is an excerpt from Harvard Medical School that further describes why we should check HRV:

"HRV is an interesting and noninvasive way to identify these ANS imbalances. If a person’s system is in more of a fight-or-flight mode, the variation between subsequent heartbeats is low. If one is in a more relaxed state, the variation between beats is high. In other words, the healthier the ANS the faster you are able to switch gears, showing more resilience and flexibility" ([2](https://www.health.harvard.edu/blog/heart-rate-variability-new-way-track-well-2017112212789)).

So HRV is not only a great tracker for overall fitness, but lifestyle as well.

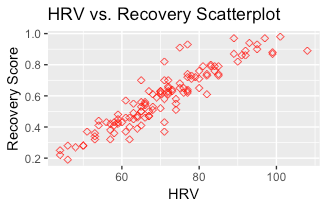


**HRV: Heart-Rate Variability**

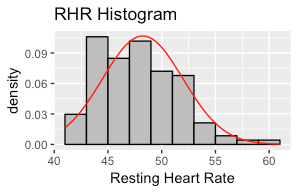
I started exploring HRV by creating a boxplot, where it showed that my HRV tended to stay somewhere between 60 and approximately 80. But what about the distribution of my HRV, were there any insights in terms of the distribution of the values?

As you can see to the left, it looks like my HRV (approximately) follows a normal distribution which I thought was pretty awesome! This histogram confirms that the values for HRV looked like they were centered around about 70.

The next thing I wanted to look into was the connection between HRV and recovery, to see if there was in fact a connection between the two. I started by looking at the correlation between the two which was 0.9198956. This lines up pretty nicely with what the research and WHOOP said about HRV being a good indicator of the body's preparedness for strain. Additionally, the visual analysis further shows the strong positive correlation between the two variables, i.e. the higher my HRV, the higher my recovery score was.



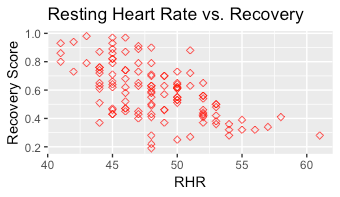
Ok so there was a reasonably strong connection between HRV and the Recovery metric, which in a way confirmed something we already knew coming in. Now let's take a look into the one of the other heart-related recovery variables: resting heart rate.



**RHR: Resting Heart-Rate**

Similar to HRV, I started with a boxplot and histogram of the values of resting heart rate.

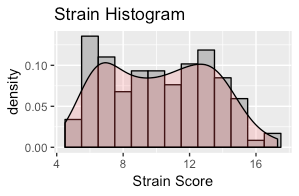
It looks like there is a slight right-skew in regard to my resting heart rate, which means the values are more condensed to the left (i.e. lower heart rates), which is a good thing! There is evidence that resting heart rate can be a good snapshot of how your heart muscle is functioning. In certain cases, according to Dr. Jason Wasfy of Harvard-affiliated Massachusetts General Hospital Heart Center, "a lower RHR can mean a higher degree of physical fitness, which is associated with reduced rates of cardiac events like heart attacks"([3](https://www.health.harvard.edu/blog/resting-heart-rate-can-reflect-current-future-health-201606179806)). So yay to no heart attacks (or at least a less likely chance of one)!



Taking a further look into resting heart rate, are there any connections between it and recovery? Let's find out. To the right is a scatter plot of the two and while not as strong as HRV and recovery, resting heart rate does appear to have a pretty solid negative correlation with recovery. This means the higher your resting heart rate, the lower your recovery score is likely to be.

**STRAIN**

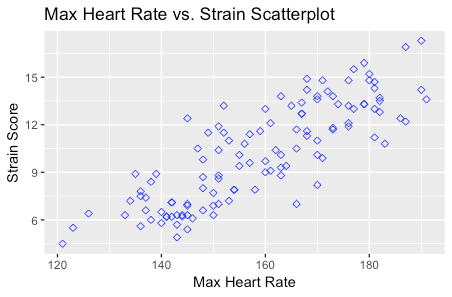
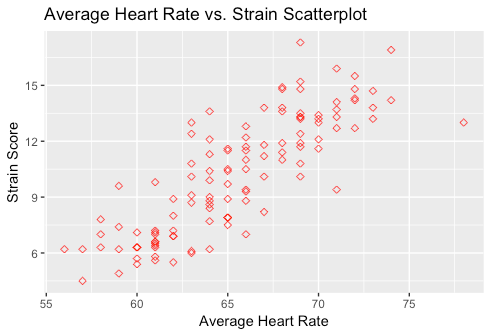
Up to this point, we've been looking largely into recovery. Let's switch gears and look into Strain, which is the variable that showcases accumulated cardiovascular load over the span of a day. This variable ranges from 0-21, with 0 being absolutely no strain and 21 being all-out, max effort.



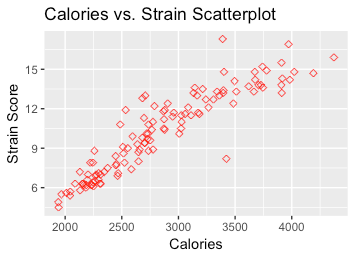
To the left is the histogram of my Strain scores for the past 4 months and a quick analysis reveals two local maxima, making it a bimodal distribution. There appears to be two groups here, with the lower centered around days which I hypothesize consist of ‘rest’ days and harder training days.

Looking at Strain in the context of heart rate, there are two variables that we have access to that may prove useful in analysis: average heart rate and max heart rate. The theory here is that the higher the average heart rate/ max heart rate, the harder you worked that day and thus the higher strain you would have. Let's take a look.

Based off the above scatterplots there appeared to be a pretty strong positive correlation between the two! When I followed that up by checking the correlation between the variables via the cor() function, R returned a correlation of 0.8253186 (averHR) and 0.8289397 (maxHR). This confirmed that there appeared to be a strong relationship between average heart rate, max heart rate and Strain. The last variable related to Strain that I wanted to take a look at was calories.



It looked like there might be a few potential outliers but overall there appeared be a strong positive correlation between calories burned and Strain which makes intuitive sense considering that the more active you are the more calories your body is going to burn.

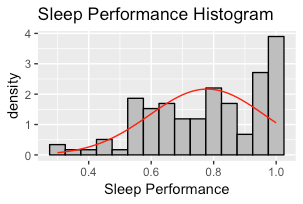


This concluded the key components of my initial analysis of Strain; I was able to gather three variables -- average heart rate, max heart rate and calories -- that I thought might prove useful when generating a model to predict the Strain metric.

**SLEEP PERFORMANCE**

The next component I wanted to take an exploratory look at was the sleep metrics, which included Sleep Performance, the times spent in the various sleep stages and sleep cycles.

Sleep Performance is a measure of the sleep the user got versus the amount of sleep that they needed. As an example, WHOOP tells the user, based off the cardiovascular load experienced that day, that they need 9 hours of sleep; if the individual gets 8 hours than the Sleep Performance score will be 88% (8 hrs. Sleep / 9 hrs. Sleep Need). So, Sleep Performance asks the question: “did you get the sleep your body needed to recover?”.

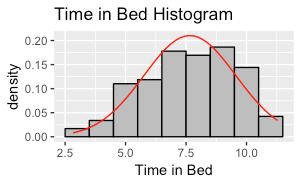


I started by looking at the distribution of my sleep performances via a histogram. I was rather surprised when I noticed that there was a tendency towards sleep performances with higher values! As you can see to the left, there is a significant spike to the right of the histogram with many of these values being between 80-100%. This surprised me because I did not think that my sleep patterns over this particular period had been that great.

This kind of contradicts what I said earlier in regard to my sleep and how I hypothesized that I wasn't getting enough based on my Strain. While we can’t rush to judgement based off a lone variable, this trend is encouraging.

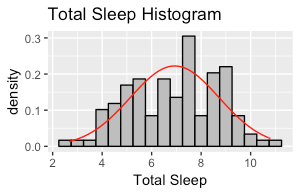
The next variable we will take a look at is Time in Bed. This variable records the amount of total time you spent in bed, from the time you pulled yourself under the covers to when you woke up and slipped your feet into your favorite pair of slippers. The relationship between time in bed and total sleep is pretty simple; the more time you spend in bed the more sleep you are likely to get.

However, this relationship is not perfect. Below is the histogram of the Total Sleep variable; intuitively, if the relationship was ‘perfect’ than we would expect the Total Sleep histogram to look similar to Time in Bed’s. This is not the case though as the Time in Bed histogram is approximately normal (with perhaps a slight left-skew) and Total Sleep appears to be a tri-modal distribution, with nodes around 5, 7 and 9 hours approximately.



This highlights the fact that sleep is not simply closing your eyes and counting sheep; it is an extremely complicated process with many things happening behind the scenes, of which many different biological factors are involved.

The sleep component of this project could probably be a project on its own, so I won’t go too far down the sleep rabbit hole. From a completely anecdotal perspective though, my sleep habits significantly improved since I started wearing the device; as a result my overall energy levels have stayed more consistent throughout the day.



It will be interesting to see what information we’ll be able to glean from the models, and if there are any valuable insights we can gather on how to better optimize one of the most important things we do as humans, sleep.

**Part 2: Creating the Models**

The primary purpose of this project is to be able to predict Strain, Sleep Performance and Recovery, which will help the user (in this case, me) to better plan out their training schedule. Additionally, a sub-goal of that was to get a better understanding of how WHOOP calculates each of these scores. The algorithms for each of these metrics is a ‘black box’; you put stuff in, it works its magic, and at the end it spits out a value. By better understanding how this black box works, we could get a better understanding of not only how they work but why we got the score that we got.

**Step 1: Cleaning the Data Set**

I wanted to include a quick section on the data wrangling process for this project. While the data set wasn’t too dirty, there were a few changes that needed to be made in order for us to be able to start creating models with it.

The first flaw was an exorbitant number of NA’s in the data; there were approximately 60 NA’s for each column. Upon further analysis there was revealed to be a sizable number of rows that didn’t have any information at all. In total we deleted approximately sixty rows with no information. Additionally, there was a missing observation for the sleep cycle variable, which we replaced with the data set’s mean (rounded to the nearest whole number) number of sleep cycles.

Next, both the Sleep Performance and Recovery variables were factor types. In the app, their values are given as a percentage out of 100 with zero being the lowest value and 100 being the highest. Using the as.numeric() function we were able to convert these values to numerics, which took a value between 0.00 and 1.00, with the former representing a score of 0% and the latter representing a value of 100%.

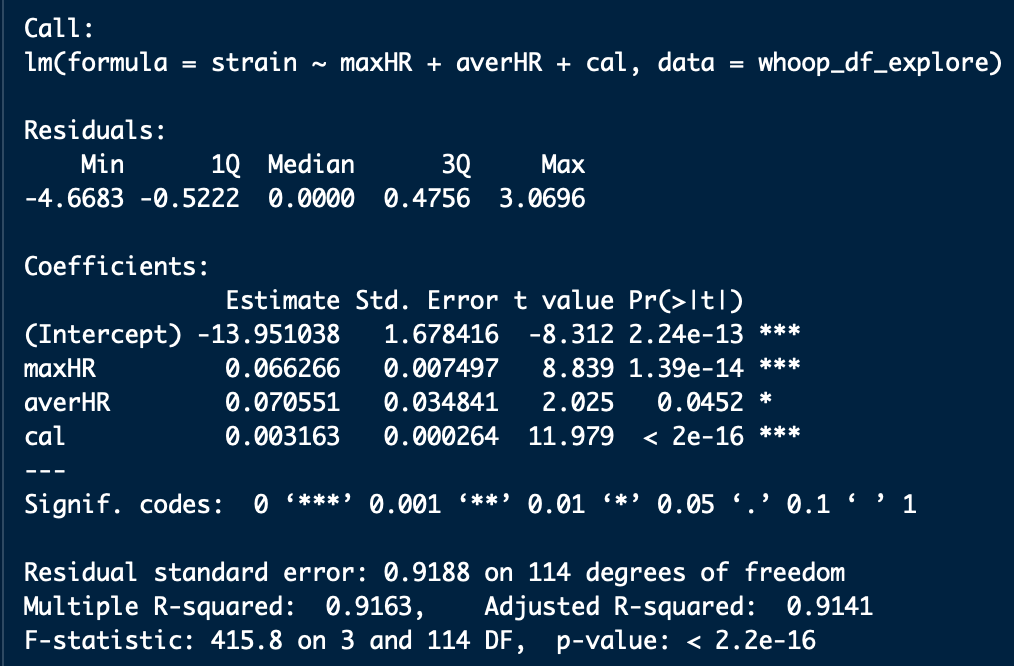
The next step was to change the variable names. There wasn't necessarily anything wrong with them however the use of uppercase, lowercase and periods is somewhat confusing and may lead to frustration as we go further with the project. So, below is the list of variables as they are represented in the updated data set (with short description):

VARIABLES IN DATA SET

* "date" (Date): day, month and year of observation
* "strain" (Strain): cardiovascular load experienced on the respective date
* "recovery" (Recovery): body's ability to adapt to training stimulus
* "sleepPerform" Sleep performance: how recovered the athlete is; largely determined by actual amount of sleep versus recommended amount of sleep need according to daily strain
* "maxHR" (Max HR): max heart rate on that respective date
* "averHR" (Average HR): average heart over the course of that entire day
* "cal" (Calories): approx. calories burned based off of cardiovascular load
* "hrv" (HRV): heart rate variability, a measure of fluctuation in the length of the time interval between successive heartbeats
* "restHR" (Resting HR): average heart rate during sleeping phase
* "timeInBed" (Time in bed): The total amount of time spent in bed, which includes time awake and time in state of sleep
* "timeLightSleep" (Light Sleep): approximate amount of time spent in light sleep stage
* "timeREMSleep" (REM Sleep): approximate amount of time spent in REM sleep stage
* "timeDeepSleep" (SWS Deep Sleep): approximate amount of time spent in SWS Deep sleep
* "totalSleep" (Total Sleep): total amount of time spent sleeping (which includes Light, REM, and SWS Deep sleep)
* "sleepCycles" (Sleep cycles): total amount of sleep cycles achieved during that day’s respective sleep session (1 stage is defined as going from Light, to REM, to SWS Deep sleep without interruption, i.e. waking up, being disturbed, etc.)

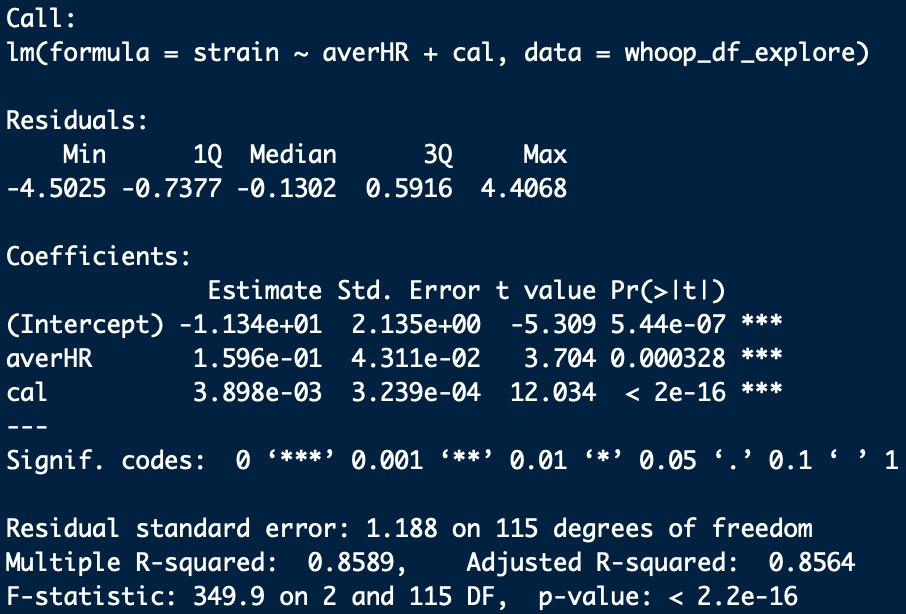
Now, the next step was to address the "date" column; the observations within that column were factors when they should represent actual dates. To do this, we loaded the lubridate package and utilized the dmy() function to change those values to a ‘Date’ type with a format of day-month-year. With our data cleaned up and ready to go, we did some initial exploratory analysis (see above) after which we moved onto creating some models!

**STRAIN MODELS**



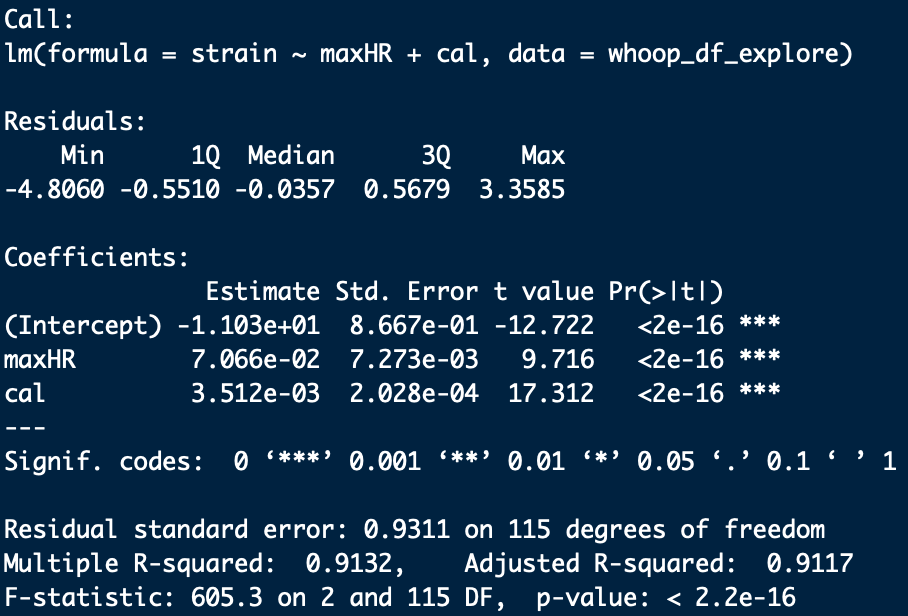
The first metric we will explore is the Strain metric. In the WHOOP app, this is one of the first metrics the user will see where it is “reported on a scale from zero to 21, [and] measures the total cardiovascular load experienced over a specified period of time - such as a workout or day - normalized such that a 21.0 represents the maximal cardiovascular load that could be attained in a day” ([4](https://www.whoop.com/the-locker/training-with-whoop-using-recovery-and-strain-to-unlock-your-potential/)).

Cardiovascular load is gauged using heart rate; every second of the day, the device is recording this. If your heart rate stays elevated for longer periods, due to events like working out or going grocery shopping with your kids, you placed more stress on your body and the result will be a higher Strain score.



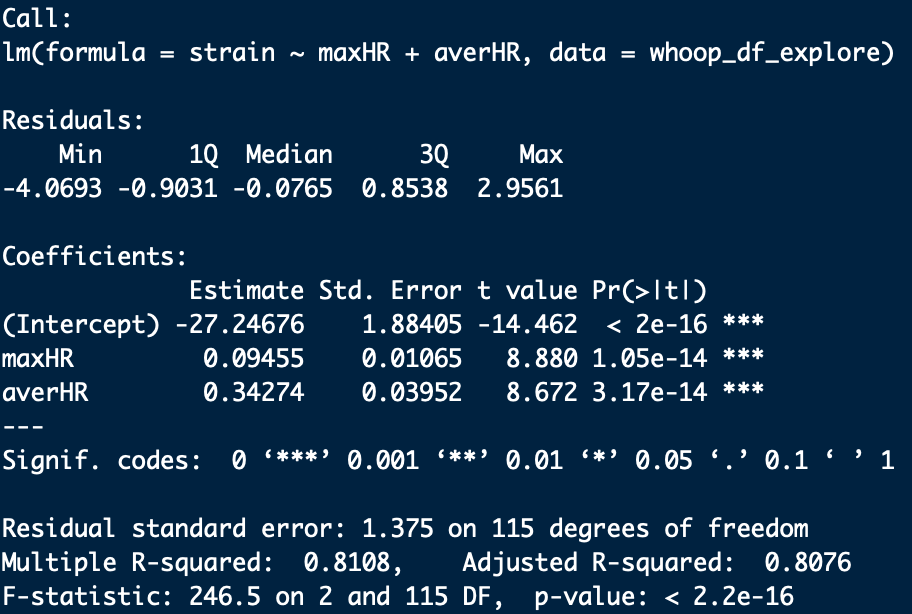
From the exploratory data analysis, we knew that three lower level variables -- max heart rate, average heart and calories -- correlated well with the Strain metric. As a result, we decided to create four models using different combinations of these variables.

The first model for Strain used max heart rate, average heart and calories as its explanatory variables; the second model used average heart rate and calories; the third model used max heart rate and calories; the fourth and final model used the two heart-related metrics, max HR and average HR. We ran the models using the lm() function and the results were encouraging.



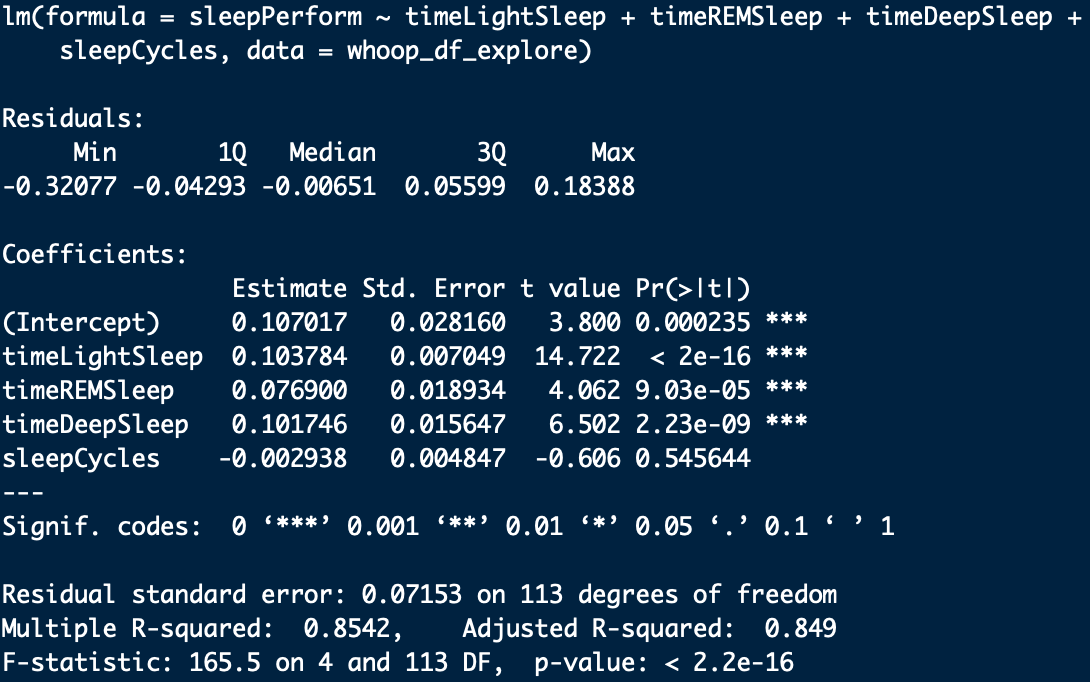
To the left are the results for each of the models as trained on the full data set. Overall, the results are a lot better than I expected! Each has a multiple R-squared value above 0.80, meaning that even in different combinations these variables appear to be strong predictors of Strain ([5](http://faculty.washington.edu/ddbrewer/s231/s231regr.htm)). However, the key word here is ‘appear’; we need to take R-squared values at face value as there are numerous precautions one should take when interpreting this number ([6](https://blog.minitab.com/blog/adventures-in-statistics-2/five-reasons-why-your-r-squared-can-be-too-high)).

However, there is one more positive that I want to point out and that is the Pr(>|t|) column. This column, which is the P value, indicates “whether the independent variable has statistically significant predictive capability. It essentially shows the probability of the coefficient being attributed to random variation. The lower the probability, the more significant the impact of the coefficient” ([7](https://infocenter.informationbuilders.com/wf80/index.jsp?topic=%2Fpubdocs%2FRStat16%2Fsource%2Ftopic41.htm)). Knowing this and taking a look at each of the variables in the models, we can see that every one of them is statistically significant meaning that the probability of the variable being due to random variation is pretty low. Again, I want to stress that we should not rush to any judgement; this is a promising start and a little later, we’ll use other methods to further validate the effectiveness the models.



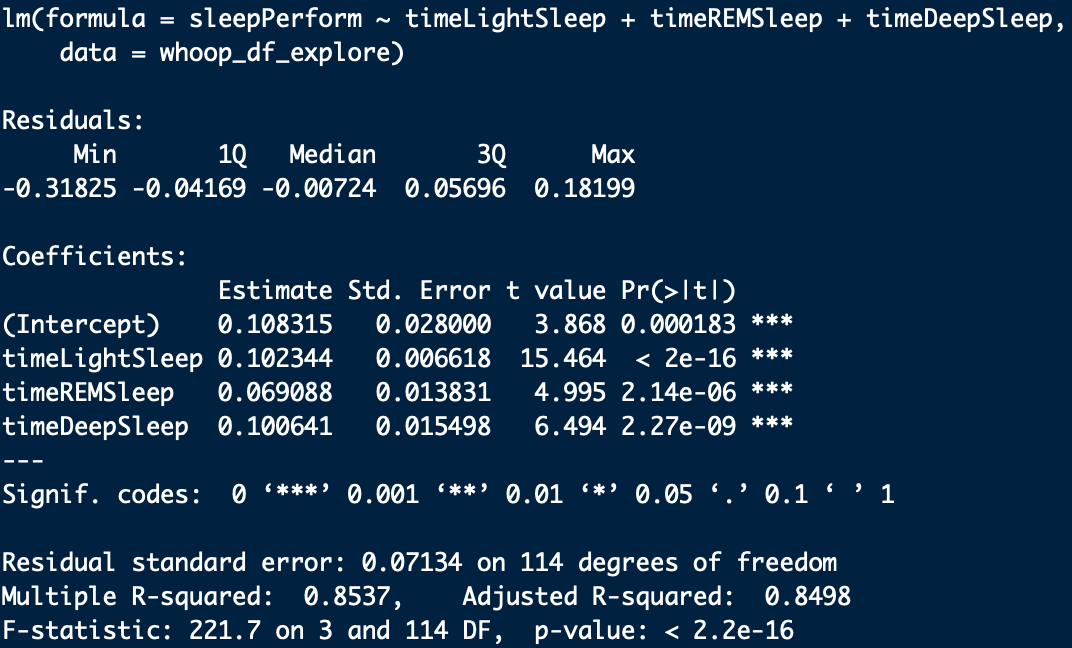
**SLEEP PERFORMANCE MODELS**

The next metric we focused on was Sleep Performance. As a high-level overview, Sleep Performance is the total amount of sleep you got divided by the amount of sleep your body needed based on that day’s strain.

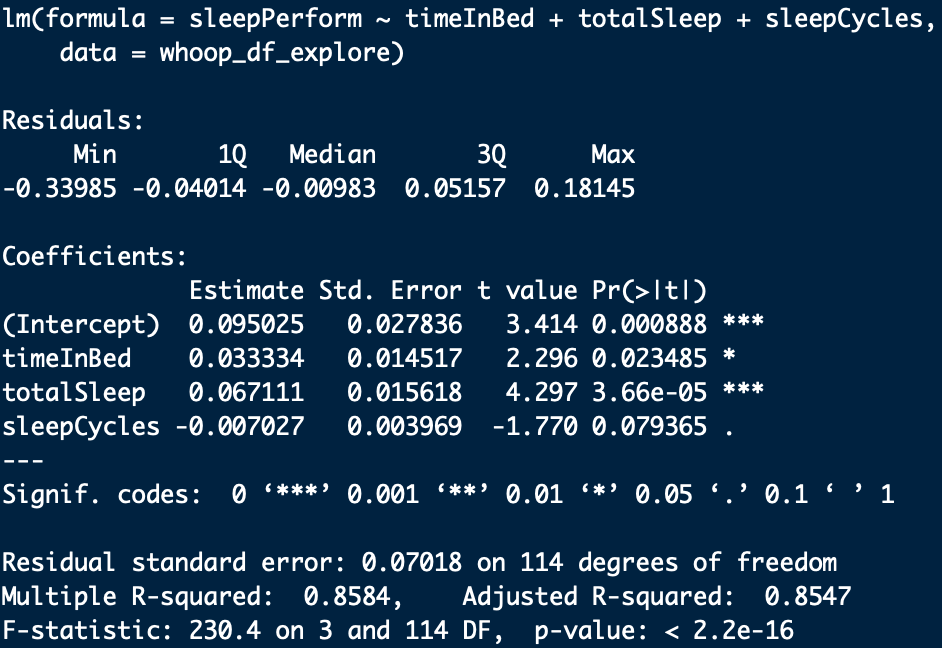


The higher your Strain metric was that day, the more stress you put on your body and as a result your body is going to need more sleep to recover. However, to look at Sleep Performance from this singular perspective is to miss out on all the interesting sleep-related variables WHOOP tracks while you catch some Z’s at night.

The following are the other sleep-related variables we had data on for this project: total time spent in bed (awake plus asleep), total time spent asleep, time spent in each stage of sleep (light, REM, and deep), and sleep cycles (successfully completing each stage of sleep without disruption).

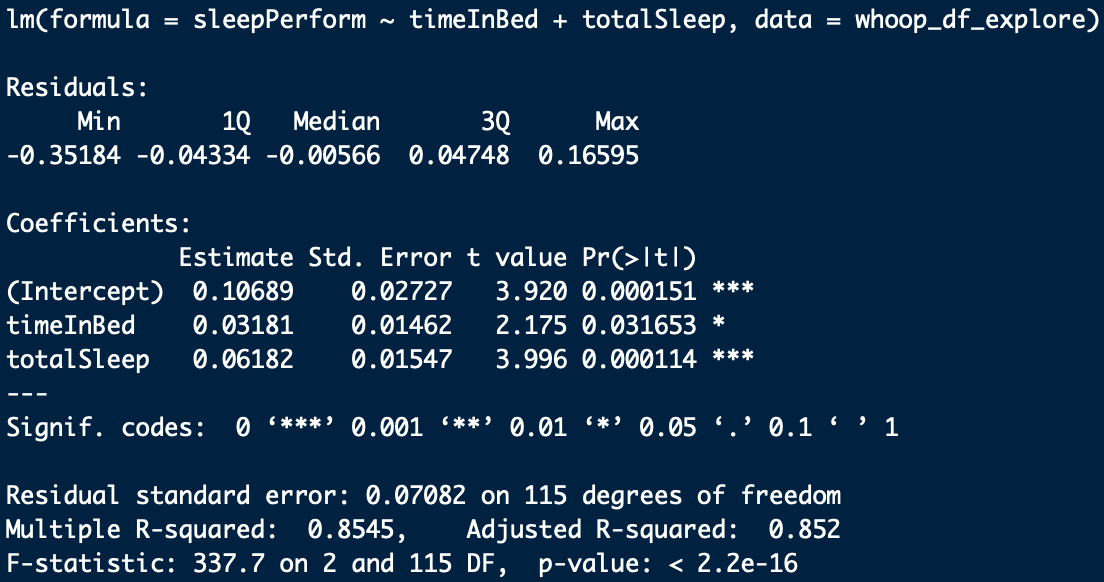


I believe that the information that the user gets about their sleep alone makes WHOOP the most beneficial wearable on the market. In recent years, numerous studies have helped us begin to understand the impact that sleep has on our health and well-being and its significant. (side note: if you are interested in learning about the importance of sleep, I highly recommend you check out UC Berkeley Neuroscientist Matthew Walker’s book, *Why We Sleep: Unlocking the Power of Sleep and Dreams*)



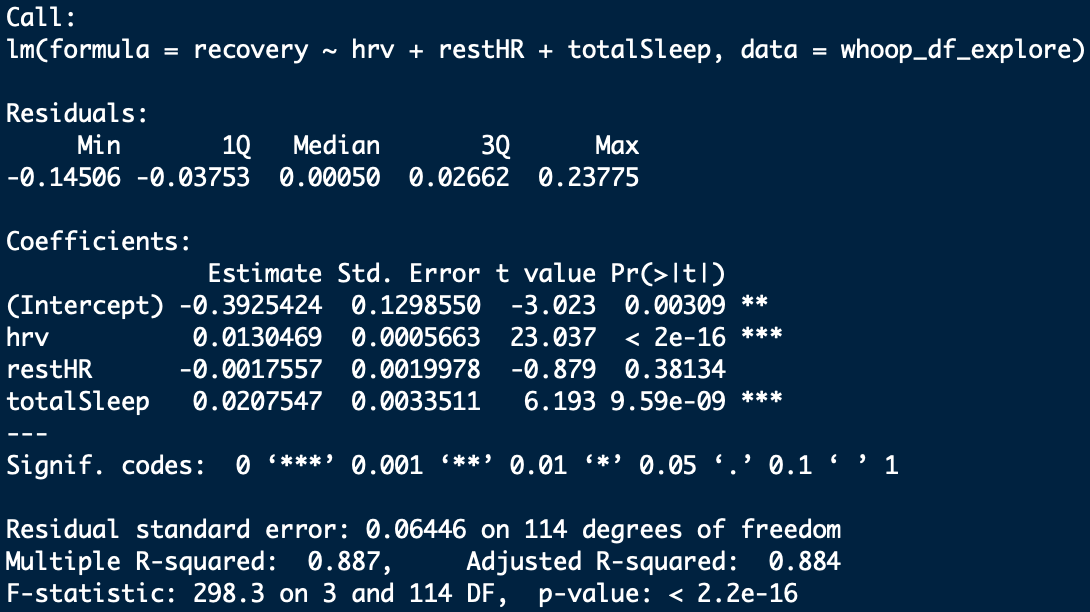
Because of the wealth of variables, I decided to take two routes when it came time to create the models: one that focused on the respective stages of sleep and one that focused on total time asleep and in bed. The R-squared is approximately 0.85 for each model, with most of the variables proving to be overwhelmingly significant. Yet, it appears that both instances where the Sleep Cycles variables is used (in models #1 and #3) it does not appear to have much of an effect compared to other models when predicting Sleep Performance. This begs the question: do we exclude it from further analysis?

At this point, no. The primary reason is there is evidence the number of sleep cycles is a critical part of a good night’s sleep ([8](https://jphysiolanthropol.biomedcentral.com/articles/10.1186/1880-6805-31-5)). Additionally, these models have been trained on the full data set and at this point we still technically don’t have validation regarding their ability to predict Sleep Performance. Since two of the models contain Sleep Cycles and two don’t, we’ll better be able to ascertain the reliability of the models once we validate them.

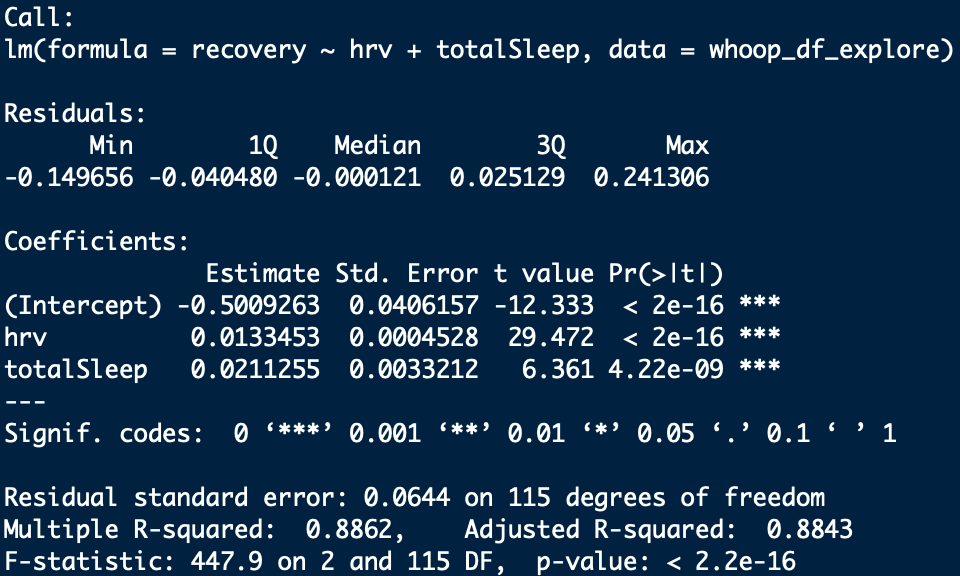


**RECOVERY MODEL**

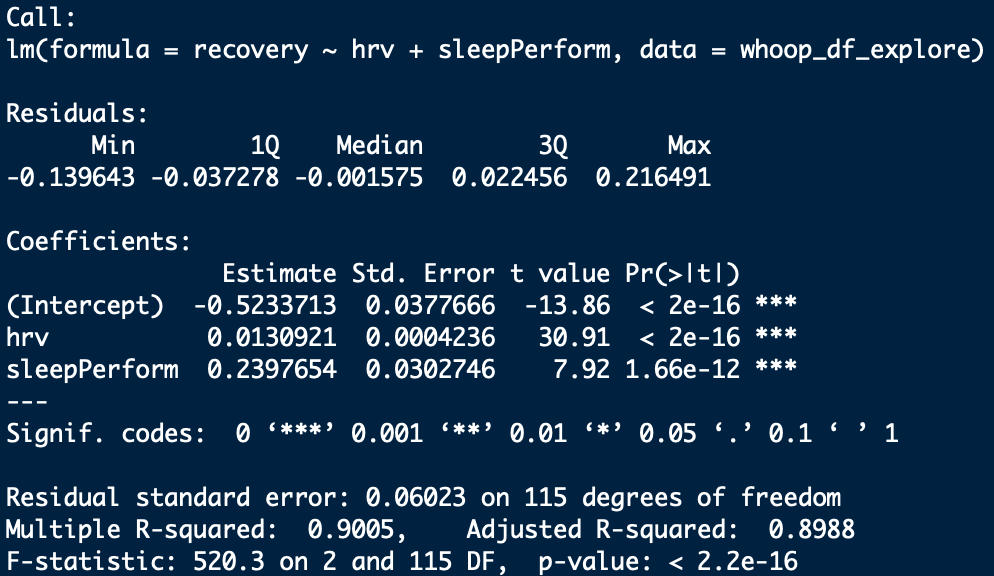
The last metric we’re going to look at is Recovery. According to WHOOP, the “Recovery [metric], reported on a scale from zero to 100 percent, measures the body’s ability to adapt to a training stimulus.” ([4](https://www.whoop.com/the-locker/training-with-whoop-using-recovery-and-strain-to-unlock-your-potential/)) In addition to the Recovery score (presented as a percentage), there are three primary variables that are associated with this particular metric in the app: HRV (heart-rate variability), RHR (resting heart-rate) and sleep (how much time you spent asleep).



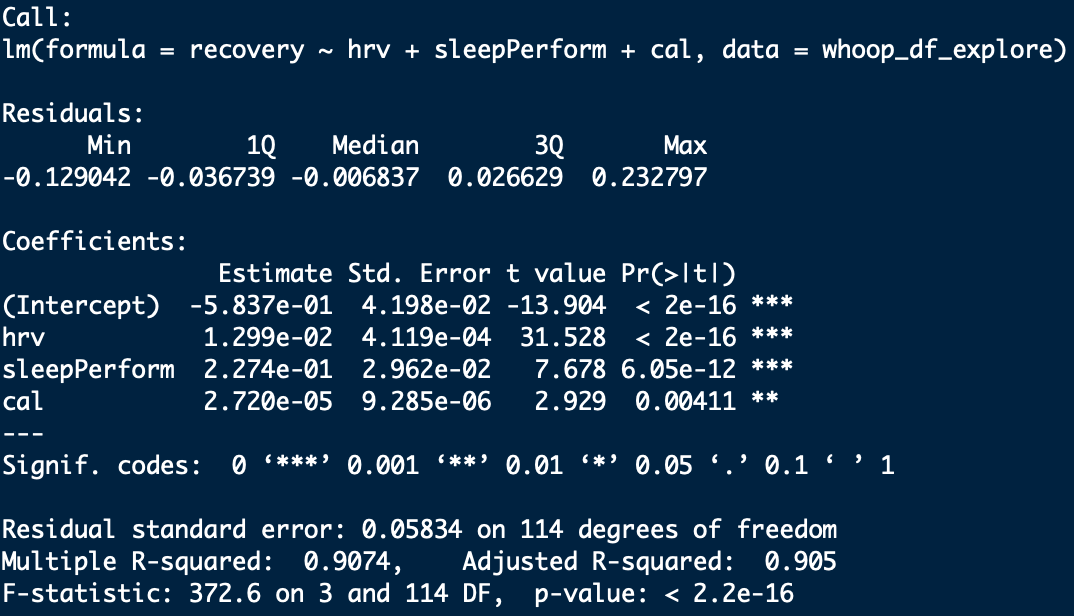
During my exploratory data analysis, the relationship Recovery had with the above variables, amongst others, was a little different compared to the other metrics. For Strain and Sleep Performance, the lower level variables associated with them had strong positive relationships. While two out of the three -- HRV and sleep -- had a positive relationship, resting heart rate had a negative relationship with Recovery. What this meant was that the higher the average resting heart rate, the lower the recovery value was going to be.



This made for a rather interesting dilemma since all three were sub-variables of Recovery: do I include resting heart rate in a model for Recovery, along with HRV and sleep? Despite the potential negative effects and since we were still in the exploratory stages of model analysis, I decided that the first model would utilize HRV, resting HR and sleep.



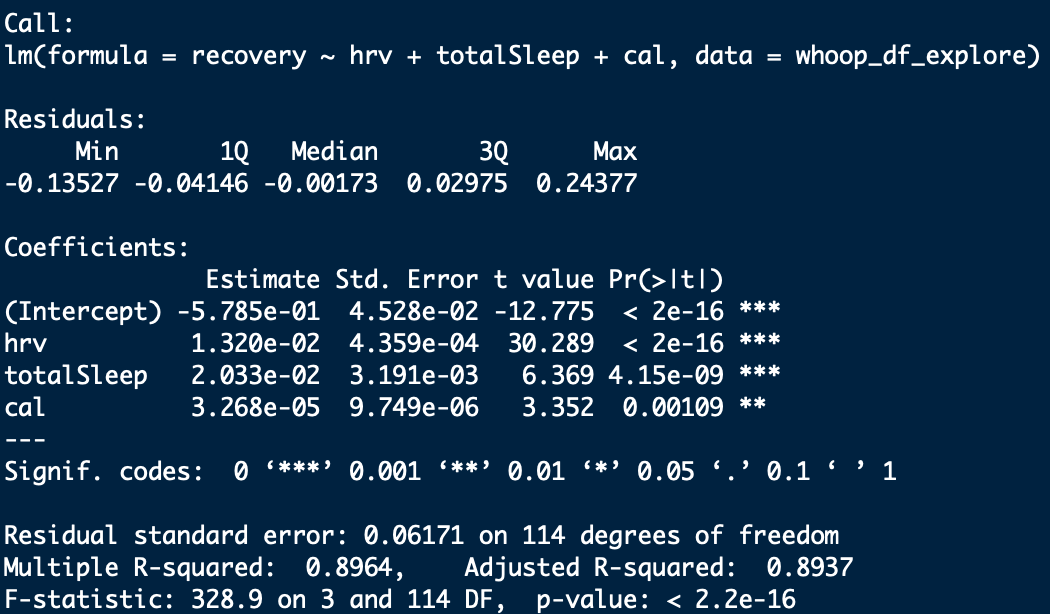
As a way to test the effects of not including resting heart rate I decided to exclude it from the second model and just use two variables, HRV and total sleep. However, my intuition was telling me that there might be a better variable to use in place of Total Sleep. Despite a somewhat ambiguous relationship from my initial analysis (and the fact that it too is a metric we’re trying to predict), I thought Sleep Performance might be a potentially better explanatory variable compared to Total Sleep. My intuition was that Sleep Performance was calculated based on how much sleep you needed versus how much sleep you actually got, so theoretically it should be a more reliable indicator of how ‘rested’ you were versus just total amount of sleep. Based off this, for the third model I decided to utilize HRV and Sleep Performance.



Another hypothesis I had was that Recovery was an all-encompassing metric; it used information from other metrics in addition to variables such as HRV to calculate its score. Intuitively this makes sense; you cannot determine recovery unless you first know how much stress the body went through, and if the body got an adequate amount of quality sleep to repair any damage done by the stress.

With this in mind, the fourth model for Recovery utilized Sleep Performance, Calories and HRV. The fifth and final model also used Calories and HRV, but Total Sleep was subbed for Sleep Performance to see if there would be any significant difference between the two.

From the initial analysis, the performance of these models for Recovery appear to be pretty good! There is some variability when it comes to their multiple R-squared score, but all are well above 0.80. Additionally, the vast majority of the variables are statistically significant. Yet, when we look at the first model, we see that resting heart rate is not significant and its P-value is pretty high. This doesn’t bode well for that particular model however we wanted to hold out excluding any model until after validating them.



**K-FOLD CROSS VALIDATION**

We have a nice set of models that show some potential for predicting their respective metrics. Yet, we now need to test them to see how good they truly are. These models were generated using the full data set which means that they’ll be good at predicting values within this particular data. What good is a model though if it can’t predict using values from outside this specific data set? Little to none. For this reason, we decided to select K-Fold Cross Validation to assess each model’s effectiveness.

The general outline of how the procedure works is as follows:

1. Shuffle the dataset randomly
2. Split the dataset into K groups (with K being equal to 10 in our case)
3. For each of these unique groups:
   1. Take the group as a hold out or test data set
   2. Take the remaining groups as the training data set
   3. Fit a model on the training set and evaluate it on the test set
   4. Retain the evaluation score (in our case the RMSE, which I’ll discuss shortly) and discard the model
4. Summarize the skill of the model using the sample of model evaluation scores ([9](https://machinelearningmastery.com/k-fold-cross-validation/))

K-Fold CV is a pretty simple process and has the advantage of generally resulting “in a less biased or less optimistic estimate of the model skill than other methods, such as a simple train/test split.” ([9](https://machinelearningmastery.com/k-fold-cross-validation/)) Choosing K (the number of folds) can be difficult as a poorly chosen value can potentially misrepresent the data. For this reason, I decided to choose K = 10, meaning that we’ll be dividing the data set into ten random folds, or groups. The primary reason for using ten was because it is very common in applied machine learning and it is “a value that has been found through experimentation to generally result in a model skill estimate with low bias and modest variance” ([9](https://machinelearningmastery.com/k-fold-cross-validation/)). Using 10-fold CV, we’ll get the evaluation score, which will be the root-mean square error (RMSE). The RMSE is an indicator of how close the observed data points are from the predicted data points based on the model. ([10](https://www.theanalysisfactor.com/assessing-the-fit-of-regression-models/)) As an example, a ‘perfect’ model would have a RMSE of 0 because the predicted values are exactly the same as the observed values.

However, perfection is nearly impossible to accomplish so in general the lower the RMSE (i.e. predicted values are closer to the observed values), the better the model is at prediction. Because K = 10, we will have 10 RMSEs per model. We’ll take the mean of these ten values and compare them to the values for the other models, with the lower values corresponding to models that more accurately predict that respective metric.

**RESULTS OF CROSS-VALIDATION & MODEL PARAMETER ESTIMATES**

*Strain*

The first model utilized all three variables -- MHR, AHR and CALS. For the full data set, the model RMSE was approximately 0.9030; what this number tells us is that the predicted Strain from the model would be somewhere between +/- 0.9 of the actual Strain. To check the validity of this model, we then ran it through the 10-fold cross validation and it returned a RMSE value of approximately 0.9358.

This was a very promising first model and set a rather high benchmark for the next models, which as it turned out, would not be beat. There were three other models tested; the second model utilized average HR and Calories; the third model utilized max HR and Calories; and the fourth and final model utilized the two heart-related variables, max HR and average HR.

However, the only other model that came close was the third model with a cross-validated RMSE of approximately 0.9454.

Additionally, the formula for the winning Strain linear regression model (i.e. model #1) was as follows:

Strain = -13.951038 + (0.066266 \* maxHR) + (0.070551 \* averHR) + (0.003163 \* cal)

This formula represents how Strain is calculated. The -13.951038 represents the linear models intercept and the numbers in front of maxHR, averHR, and cal are the model’s parameter estimates for their respective variables. For example, let’s plug in a 180 for maxHR, 70 for averHR, and 3,500 for cal.

MaxHR: 0.066266 \* (180) = 11.92788

AverHR: 0.070551 \* (70) = 4.93857

Cal: 0.003163 \* (3500) = 11.0705

With these example inputs, the model would predict a Strain score of approximately 13.9.

Strain: -13.951038 + 11.92788 + 4.93857 + 11.0705 = (~)13.9

*Sleep Performance*

Compared to the Strain models, the RMSE of the cross-validated models were extremely narrow with each being approximately +/- 0.07, or 7%, of the actual Sleep Performance value (which is reported as a percentage out of 100 in the app). We had to go out to the ten-thousandths place, to determine the ‘winning’ model which ended up being the third one which used the variables for time in bed and asleep plus total sleep cycles.

The third model’s cross-validated RMSE was 0.07147 (7.147%), which barely beat out the fourth model’s cross-validated RMSE of 0.07179 (7.179%). Despite the third one technically having the lowest RMSE value, the closeness of the four Sleep Performance models was interesting. For further research it would be interesting to dig down a little deeper to determine why and how these different variables were able to produce such similar numbers.

Additionally, the formula for the winning Sleep Performance linear regression model (i.e. model #3) was as follows:

Sleep Performance = 0.095025 + (0.033334 \* timeInBed) + (0.067111 \* totalSleep) +

(-0.007027 \* sleepCycles)

This formula represents how Sleep Performance is calculated. The 0.095025 represents the linear models intercept and numbers in front of timeInBed, totalSleep, and sleepCycles are the model’s parameter estimates for their respective variables. For example, let’s plug in 8.5 (hrs.) for timeInBed, 8.0 (hrs.) for totalSleep and 5 for sleepCycles.

timeInBed: 0.033334 \* (8.5) = 0.283339

totalSleep: 0.067111 \* (8.0) = 0.536888

sleepCycles: -0.007027 \* (5) = -0.035135

With these example inputs, the model would predict a Sleep Performance score of 0.88, which represents 88%.

Sleep Performance: 0.095025 + 0.283339 + 0.536888 + (-0.035135) = (~)0.88

*Recovery*

When cross-validated, the Recovery models’ results were also extremely narrow with the RMSE being approximately +/- 0.06, or 6%, of the actual Recovery value. However, there was one that slide in right under 0.06, and that was the fourth model with a RMSE of approximately 0.05987.

Remember that this model utilized both Calories and Sleep Performance in addition to HRV to calculate the value of Recovery. The cool thing about this result was that opens the door for further investigation into Recovery being an all-encompassing metric.

Additionally, the formula for the winning Recovery linear regression model (i.e. model #4) was as follows:

Recovery = -0.5837 + (0.01299 \* hrv) + (0.2274 \* sleepPerform) + (0.00002720 \* cal)

This formula represents how Recovery is calculated. The -0.5837 represents the linear models intercept and the numbers in front of hrv, sleepPerform and cal are the model’s parameter estimates for their respective variables. For example, let’s plug in 100 for hrv, 0.80 for sleepPerform and 3,500 for cal.

hrv: 0.01299 \* (100) = 1.299

sleepPerform: 0.2274 \* (0.80) = 0.18192

cal: 0.00002720 \* (3500) = 0.0952

With these example inputs, the model would predict a Recovery score of approximate 0.99, which represents 99%.

**CONCLUSIONS & RECOMMENDATIONS**

*Conclusions*

1. For Strain, the first model that utilized max heart rate, average heart rate and calories will be able to predict a user’s daily Strain Score within approximately +/- 0.93 of the actual value.
2. For Sleep Performance, the first model -- with time spent in the light sleep stage, REM stage, deep sleep stage and sleep cycles as explanatory variables -- performed the best and will be able to predict a user’s sleep score within approximately +/- 7% of the actual value.
3. For Recovery, the fourth model -- which utilized HRV, Sleep Performance and calories as explanatory variables -- performed the best at predicting the Recovery metric, being within approximately +/- 6% of the actual value.

*Recommendations for User*

1. By using the Strain model, not only will it help assess how hard you should push your body on any given day, but it also has the additional capability to help with diet. Since calories are one of the variables, you can use the estimated number of calories to gauge meals and better plan when and what to eat.
2. The Sleep Performance model can help individuals prioritize sleep. Sleep is a crucial component of recovering from a day’s stressors and is an often-overlooked aspect of training. Understanding how much time you need to spend in each stage (thus allowing you to see how much sleep you need overall) will force the user to prioritize pre-bed routines and activities that promote restfulness.
3. The Recovery model can be used to determine which days of the week the user wants to train harder or easier. Knowing they’re going to have high-recovery score can help them plan for more time training on that day; additionally, when the user has a full schedule, planning a ‘recover’ day will allow users to maximize their overall training regimen.

*Recommendations for Further Research*

1. Include data on circadian rhythms. One of the biggest influences of sleep is our circadian rhythm, which is a 24-hour cycle governing our energy levels. There is evidence that this plays a large role in determining our sleep quality. Having data on specific times when a user goes to sleep and wakes up in the morning may help increase the robustness and/or accuracy of the Sleep Performance model.
2. Take a deeper dive into HRV. There is a body of evidence that suggest HRV has a lot of potential as an indicator of how prepared the body is to take on strain, hence WHOOP using it as a low-level variable associated with the Recovery metric. However, this project did not look into it on a singular basis. It would be interesting to find if there are any ways you could predict HRV, as it is a large component of Recovery.
3. Time spent in different heart rate zones. Admittedly calculating average heart rate is complicated. It is not realistic or feasible even to expect users to calculate their heart for every second of the day. What would be useful though would be the ability to tell users how much time they could spend in certain heart rate zones. They could then design their training plan for that day according to how high they could get their heart rate and for how long it should be within that zone. This has the potential to prevent overtraining.

***Sources***

1. <https://www.whoop.com/the-locker/an-athletes-guide-to-hrv/?gclid=Cj0KCQjw3KzdBRDWARIsAIJ8TMRAgo8EOPlnujmFKGPoqavNn75jrJfOYJOKkG4aV5R9IpZmPOcDkwwaAnxiEALw_wcB>
2. <https://www.health.harvard.edu/blog/heart-rate-variability-new-way-track-well-2017112212789>
3. <https://www.health.harvard.edu/blog/resting-heart-rate-can-reflect-current-future-health-201606179806>
4. <https://www.whoop.com/the-locker/training-with-whoop-using-recovery-and-strain-to-unlock-your-potential/>
5. <http://faculty.washington.edu/ddbrewer/s231/s231regr.htm>
6. <https://blog.minitab.com/blog/adventures-in-statistics-2/five-reasons-why-your-r-squared-can-be-too-high>
7. <https://infocenter.informationbuilders.com/wf80/index.jsp?topic=%2Fpubdocs%2FRStat16%2Fsource%2Ftopic41.htm>
8. <https://jphysiolanthropol.biomedcentral.com/articles/10.1186/1880-6805-31-5>
9. <https://machinelearningmastery.com/k-fold-cross-validation/>
10. <https://www.theanalysisfactor.com/assessing-the-fit-of-regression-models/>