Analyzing Sleep Through Smart Technology: A Case Study of Will Foote's Fitbit Data

Kyle Fang and William Foote

I. Abstract

A major component to a college student's performance in school is sleep. We know that especially in college, the amount of time a student has for sleep is limited, we sought to understand which factors other than time contribute to high quality sleep.

In early August 2019, Fitbit introduced the "Sleep Score" metric to quantify one's sleep quality. We gathered data from Will Foote's Fitbit tracker, which measures different aspects of his daily activities as well as different aspects of his sleep every night (e.g. minutes asleep, minutes in deep sleep, number of times he wakes up in the middle of the night, etc.).

Though the raw Fitbit data spanned August 2019 to May 2020, some of the data points were missing for various weeks. Thus we chose to look at the data from Fall Quarter 2019, because it has minimal missing points (12 nights of sleep not tracked over 76 days studied). In addition, we realized that our results would be more useful if we studied a time range in which the studied sleep occurs under mostly uniform conditions (e.g. location slept, daily schedule, etc.).

With our data selected, we sought to answer three main questions. What variables play a role in Will's quality of sleep that can be modified via lifestyle changes to bring about better sleep? What other uncontrollable factors affect Will's sleep? Do these factors support previous literature on the science of sleep or refute it?

II. Overview of data

Over the past year or so, Will has been monitoring his sleep using the sleep tracker on his fitbit. Every day, the data was transmitted from the fitbit to Will's online account, and we were able to download it from his account's database. We collected this data because we wanted to see how different predictor variables (heart rate, hours of sleep, breathing, etc.) affected his quality of sleep, quantified by the sleep score number. The sleep score ranges from 1 to 100, with 100 being the best quality sleep, and 1 being the lowest quality sleep.

Since we wanted an accurate measure of Will's normal sleep, we looked at data ranging over the course of a regular UCLA school quarter. Specifically, we looked at his sleep during Fall Quarter 2019 - 9/30/2019 (Monday, Week 1) to 12/13/19 (Will's last final). We did not want to look at dates outside this range because his sleep schedule during break is drastically different from his sleep schedule during school.

III. Results

Our final model:

$$(OverallScore)^3 = \beta_0 + \beta_1 NumberOfAwakenings + \beta_2 Restlessness + \beta_3 StartTime + \beta_4 EndTime + \beta_5 ProportionOfREMSleep$$

Where

- $\beta_0 = 183989.8$
- $\beta_1 = 5426.4$
- $\beta_2 = -2196857$
- $\beta_3 = -537297.9$
- $\beta_4 = 563825.9$
- $\beta_5 = 878899.9$

- *NumberOfAwakenings = The number of times Will woke up during the night as registered by the fitbit*
- *Restlessness = the score assigned to Will's restlessness for a given night*
- *StartTime = the time Will went to bed in units of proportion of a day after 6 p.m.*
- *EndTime = the time Will woke up in units of proportion of a day after 6 p.m.*
- *ProportionOfREMSleep = minutes of REM sleep divided by minutes of total sleep*

As you can see in our summary table below, all the t statistics have p values smaller than the 0.05 significance level. This tells us that each predictor variable has a significant role in explaining variation in the response variable, given that all the other variables are already in the model.

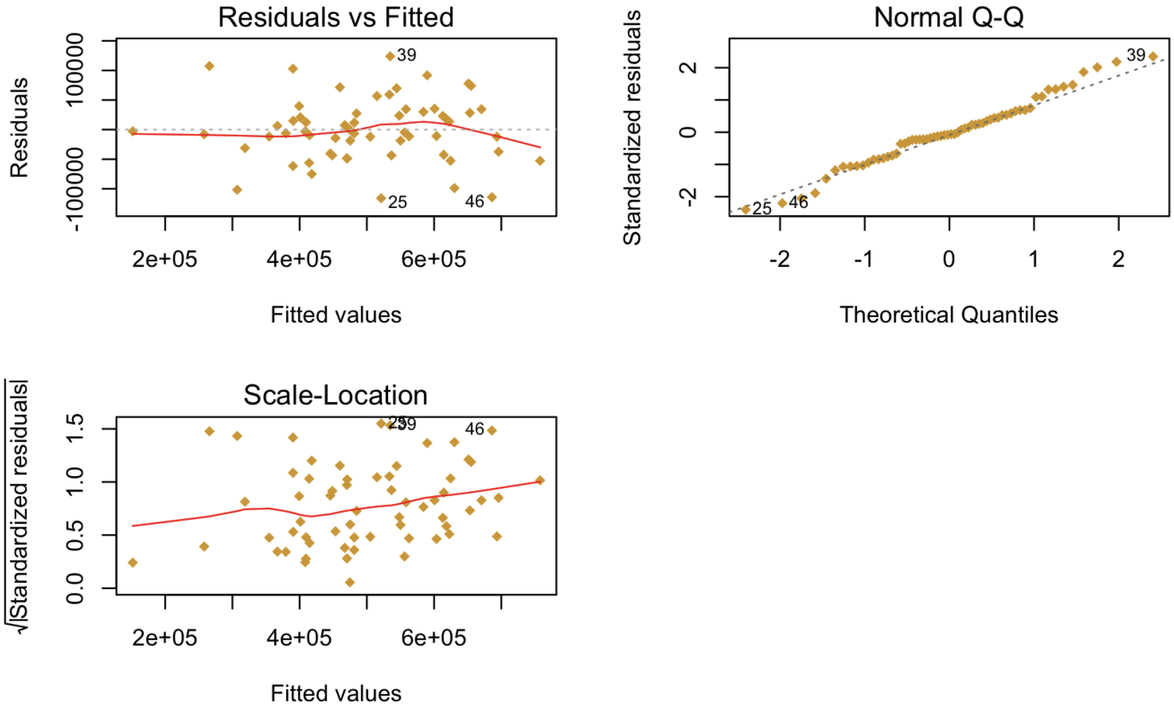
```
Call:
lm(formula = overall_score ~ . - Date, data = sleep_final)

Residuals:
    Min      1Q  Median      3Q     Max
-116016  -35858   -3568   28015  123789

Coefficients:
                      Estimate Std. Error t value Pr(>|t|)
(Intercept)           183989.8    68036.5   2.704  0.00905 **
Number.of.Awakenings    5426.4      980.2   5.536 8.53e-07 ***
restlessness        -2196857.0   311923.1  -7.043 2.94e-09 ***
start_asTime         -537297.9   157634.0  -3.409  0.00122 **
end_asTime            563825.9   125640.5   4.488 3.63e-05 ***
prob_rem              878899.9   147856.0   5.944 1.87e-07 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 54330 on 56 degrees of freedom
  (14 observations deleted due to missingness)
Multiple R-squared:  0.8444,    Adjusted R-squared:  0.8305
F-statistic: 60.76 on 5 and 56 DF,  p-value: < 2.2e-16
```

Summary table

In the residual plot (below), we conclude that there is no trend in the smoother line. There is a dip at the end, but we believe that is because there are only 3 points after 8e+05 and all three of those points have negative residuals. If those three outliers were not there, then we think the line would have had no trend at all.

In the QQ plot, most of the points fall on the smoother line, so the residuals seem to follow normal distribution.

Lastly, the scale location plot seems to have a bit of an overall trend. It’s not clear from the scale-location plot itself whether the upward motion of the line is a trend or a result of a very small number of points near the end. In the residual plot, there does not seem to be an obvious fan shape in the residuals, so it does not seem like the constant variance aspect of the model is nullified. Overall, we conclude that the constant variance condition is still held, but it is a weakness of our model.



Diagnostic Plots

IV. Discussion

This topic was interesting to us because we recognize the importance of sleep, but at the same time, we as college students know and experience having limited time for sleep due to the busyness of college life. We know that the amount of sleep we have at night plays a role in how energized we are the following day, but we were interested to see what other factors contribute to a restful sleep. Especially as busy college students, we often cannot control how much time we can sleep at night. If we can find the other factors that affect sleep quality and if these factors are things we can control, then this information would be very helpful to us and allow us to change our lifestyles to have more restful sleep, allowing us to perform better in school.

Will's fitbit account provided 2 files of data -- a sleep file containing information specific to his sleep (the sleep score, heart rate, restfulness, etc.) and an activities file containing broader categories of Will's activity throughout the day and night (calories burned during the day, hours he was sedentary, sleep duration, etc.). Variables from both of these files were useful so we combined them into one data frame.

We faced some challenges in preparing the data. For instance, Will had data from the daily activities data set for every day in the time period but not for every day in the time period for the sleep data set. There were a couple days where Will did not wear his fitbit when he slept, so we had to reformat and match the dates of these data frames before concatenating them together. We filled in missing data with NA values.

In addition, we wanted to see if the time Will went to bed and time he woke up were significant, so we had to reformat the time into military time and convert it into a number that represented time relative to a fixed reference point. This process proved to be more difficult and tedious than we had anticipated.

Another challenge was naps. We realized that there were occasional days when Will napped, and for certain naps, an additional separate sleep score was assigned to the naps on those days. In total, however, there were only 5 naps long enough to generate a sleep score in the 3 month period, so we manually deleted the nap sleep scores as we were interested in what factors contribute to the quality of a full night's rest and not to the quality of a short midday nap.

Looking at the sleep score data, we noticed that it was left skewed, so we predicted that we would need to transform the response variable by exponentiating it to some power. We employed the Box Cox method to see what exact transformation to apply and were recommended to transform this response variable by exponentiating it to a power of 3.

We looked at histograms of the predictor variables, and they all seemed to be normally distributed. To confirm this, we used the Box Cox method to see if it suggested transforming any of the predictor variables. As we anticipated, the boxcox did not suggest any transformation, so we kept the predictor variables untransformed.

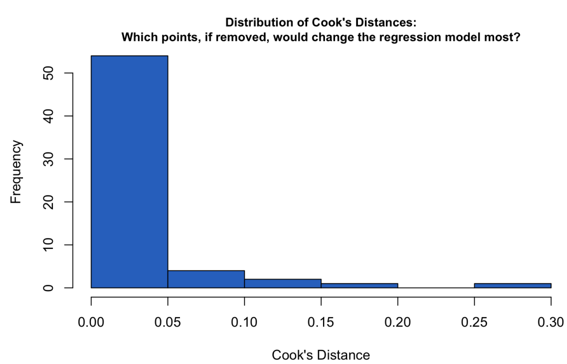
In order to determine which variables belong to the model and which do not, we first looked at all the predictor variables and tossed out two categories of variables: irrelevant variables and variables that we knew would be collinear with other variables. Irrelevant variables included the dates, sleepscore id, number of floors climbed that day, and many others. Variables that were collinear included the pair of minutes asleep and minutes awake variables (when Will was in bed), the group of minutes of REM sleep, minutes of light sleep, and minutes of deep sleep, and other groups. For the group of collinear variables, we only kept one variable out of these since the number of minutes asleep is directly proportional to the number of minutes awake. For the more complicated groups like minutes of REM sleep, minutes of light sleep, and minutes of deep sleep, we experimented with keeping two in, and then just keeping one, looking to see which individual or group of variables were significant and not collinear.

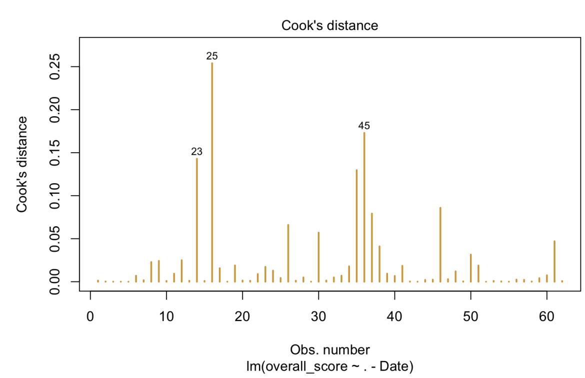
After picking out these variables, we experimented with models containing the rest of the variables. After this point of our process, a high p value was the factor that eliminated further variables. For each model we tried, we double checked the variance inflation factor to ensure that collinearity was not the reason p values for certain variables were too high. After playing around with different models with different variables, we concluded that there are 5 variables that have a significant role in explaining the overall sleep score.

Overall, we found three bad leverage points. All of these three bad leverage points were also high influence points, so it was important for us to understand why these points were outliers. All three of these data points are noticeably different in the amount of sleep and the restlessness score. Throughout Fall quarter, Will slept an average of 7.28 hours with a standard deviation of 1.85. During these bad leverage point dates, however, he slept 4.82 hours, 6.23 hours, and 10.5 hours. The hours alone suggest that his sleep those nights would be irregular.

His restless scores over the time period, which have an average of 0.11245 and a standard deviation of 0.0232, were 0.1566, 0.1794, and 0.1486. These scores are certainly outliers for this predictor because they constituted 3 of the top 4 restlessness scores.

All three of these data points had sleep scores that were lower than the mean. This makes sense to us because having an anomalous quantity of sleep (either too much or too little) as well as having a very restless night would logically translate to low quality of sleep. We trust that the fitbit measured these parameters on these dates accurately and that it gave an accurate sleep score for those nights, so these bad leverage and influence points do not alter our conclusion.





(left) *A histogram of the Cook's Distances (where high values indicate the regression model would be affected if they were removed, and could be a sign the model is not robust) shows there are a few points that have much higher influence than most.*
(right) *Investigating further, three points, observations 23, 25, and 45, have the highest Cook's Distances and have the highest influence on the regression model.*

We were not surprised by most of the predictor values. What the predictor values from our model tell us is that high restlessness in sleep is correlated with a low quality of sleep. A late bedtime is correlated with a low quality of sleep. A late wakeup time is correlated with a high quality of sleep. A higher proportion of REM sleep is correlated with a high quality of sleep. These four conclusions make intuitive sense.

However, the model states that a high number of awakenings in the middle of the night is correlated to a high quality of sleep. This is surprising to us because we would expect that the smaller the number of times Will wakes up in the middle of the night, the better his sleep. One thing to note, however, is that while the coefficients for the other variables in the model range from 100,000 to 2,400,000, the coefficient of the Number.of.Awakenings variable is only 5426. This suggests that Number.Of.Awakenings plays a really small role in explaining the sleep score compared to the other variables in the model.

Now this small slope may be small due in part to the relatively large range of the Number.Of.Awakenings variable compared to the other 4 variables. Number.of.Awakenings has a range of 30, while the other variables have ranges less than 0.3. Nonetheless, this accounts for only a small part of the gap in slopes between Number.Of.Awakenings and the other predictors,

so Number.Of.Awakenings really does have a very small effect on sleep score. Since the overall effect of Number.Of.Awakenings on sleep score is small, an unexpected positive slope for this variable is not too problematic to our interpretation of the model.

V. Conclusion

Considering the data we had to analyze, we believe we answered our main questions at hand. Even though our study is purely observational, limiting our ability to make claims about what *causes* good or bad sleep, in terms of understanding what can be changed in Will's life to make for a better night's sleep, our research suggests going to sleep earlier and waking up later; more sleep is typically better sleep.

Though the other three factors we found to have a significant correlation with sleep score cannot easily be modified by lifestyle changes, our findings are still useful in helping us understand the science of sleep better and corroborating existing research on this topic. Most complexly, we found that a greater number of awakenings during a night correlated with a better sleep score. This surprising finding might be explained in two ways. For one, studies show that around 4-5 cycles of sleep are important for getting a good night's sleep (Crane, 2016), and a certain number of awakenings between cycles is a sign of normal, healthy sleep (Moline et. al, 2014). However, awakenings as a result of poor sleeping conditions or external factors like physical or mental stress can be signs or causes of a bad night's sleep. Perhaps in further research, analyzing the number of awakenings against a baseline value might change our results. Matching our intuition, higher restlessness values were typically associated with lower sleep scores.

Our research corroborates Fitbit's explanations about their sleep scores. Our research found that higher proportions of REM sleep out of total sleep correlated with higher sleep scores, and as Fitbit's article on sleep scores explains, "How much time you spent in deep and REM sleep—the more time you spend in these sleep stages, the better your score" ("What's sleep score in Fitbit app?", 2020). While this could be a result of how Fitbit calculates its sleep score (i.e. in a way, we've just derived part of their sleep score formula), previous research has also suggested that a certain amount of REM sleep is important to get quality sleep.

Our research should not be considered as the final step for us or this topic in the slightest. Our findings are certainly limited, because of its observational (versus experimental) nature, and more importantly because our findings are only as good as the assumptions Fitbit considered when developing the sleep score metric. That is, is Fitbit's sleep score really a good indicator of the quality of sleep?

Also, our study was limited by time and available data. If we had more time, looking for relationships between the number of sleep cycles in a night, the sleep cycle Will woke up in, estimated oxygen variation during sleep, and levels of stress for a given day, could all enhance our analysis. Also, one might consider a metric like variation in sleep start time from night to night, as the National Institute of Neurological Disorders and Stroke argues that maintaining a sleep schedule can benefit one's quality of sleep. With students sleep-deprived around the country, more must be done so that sleep might be a tool for combating stress instead of a lack thereof being the cause of it.

VI. Data and References

Our data:
https://docs.google.com/spreadsheets/d/1wlW12e-8t-EUfq43eAYj-z9zS9uuz17MWpylSg4B1BY/edit?usp=sharing

References:
"Brain Basics: Understanding Sleep". *NIH*. 2019 Aug 13
https://www.ninds.nih.gov/disorders/patient-caregiver-education/understanding-sleep#2

Crane, R. "Enhance Your Sleep Cycles". *Alaska Sleep Clinic*. 2016 Sept 30.
https://www.alaskasleep.com/blog/enhance-your-sleep-cycles#:~:text=Each stage%0cumulates to REM,9 total hours of sleep.

Moline, M., DiBonaventura, M., Shah, D., Ben-Joseph, R. "Impact of middle-of-the-night awakenings on health status, activity impairment, and costs". *NCBI*. 2014 Jul 23
https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4114905/

"What's sleep score in the Fitbit app?" *fitbit*. 2020 Apr 6.
https://help.fitbit.com/articles/en_US/Help_article/2439