

Homework 3
Problem 4
Ayad Masud
733009045

I certify that I have personally done the coding, generated the figures and written the report without aid from anybody else, and that I have not plagiarized, self-plagiarized, or used AI-generated text. I certify that I have acknowledged any sources I used to complete this assignment. ARM.

1 Part 1: Comments Received in Homework 2

Comment: “Your hypothesis is vague, and you are confounding time and intensity. If you compare "intense workout at night" vs light exercise" in the evening, you will not be able to determine whether it was time or intensity that affected the sleep. There are also inconsistencies in the self-report measures. At the start of the report, you mention using the Karolinska instrument, and in section 4 you state that you will use the Pittsburgh instrument. Your preliminary results are based on a loosely measured independent variable (easy/moderate/intense). I would suggest you control the independent variable. For example, choose an easy run (e.g., 8 min/mile for 2 miles would have been easy for me in my prime) vs. an intense run (e.g., 6 min/mile for 2 miles would have been hard for me). Then alternate, half the days you do the run in the morning, and the other half in the evenings. That may improve the odds you get some interesting results.”

How I fixed it: After looking at my experiment, I now understand that I cannot measure both time and intensity to see how it affects my quality of sleep at the same time. I won't be able to tell which variable is responsible for my sleep. Therefore, I have decided to measure just the intensity variable and how that affects my sleep. This is because I tend to workout the same time of day anyway and only changed what time I work out for this experiment. Just measuring the intensity variable aligns more with how I go about my daily life and what really affects my sleep quality. I will measure the levels of intensity by using heartrate thresholds and a heartrate monitor. For example, light, aerobic, and anaerobic zones. As for the inconsistencies in the self-report measure that is because I found out about the Karolinska instrument after finishing the homework 2 report and found that was better for my experiment than the Pittsburgh instrument. It was a silly mistake that I forgot to change it to the Karolinska instrument everywhere on my report. I have used only the Karolinska instrument for my self-reports.

2 Part 2: Data Analysis

Figure 1 is a correlation matrix between the heart rate, sleep quality, and sleep time in the dataset. We can see that the only meaningful correlation is between heart rate and sleep quality. The rest of the features are not correlated at all or are only slightly correlated. For example, the sleep time and the heart rate almost have zero correlation, while sleep time and sleep quality have a slight negative correlation, but it's so small that's almost zero correlation as well. Figure 2 shows box plots of each heart rate zone and the average quality of sleep that resulted in each of those zones. We can see that most of my exercise took place in the moderate range between 120 bpm and 160 bpm. We can also see an upward trend. With higher heart rate efforts, the sleep quality becomes better. Figure 3 is a plot showing heart rate plotted against sleep quality. In this plot we cannot see the positive correlation between the two features as well as we could see it in the correlation matrix and the box plots. The slope is 0.0164 which indicates that to some degree as the heart rate increases the quality of sleep also increases. Let us define a null hypothesis and an alternative hypothesis to conduct a hypothesis test. The null hypothesis is that heart rate has no significant effect on sleep quality. The alternative hypothesis is the thing we are trying to prove. Therefore, the alternative hypothesis is that the heart rate has a significant effect on sleep quality. Running the hypothesis test using python statistic functions we get a p-value of 0.00109. This is less than the 0.05 threshold for significance. This shows that there is a statistically significant relationship between heart rate and sleep quality.

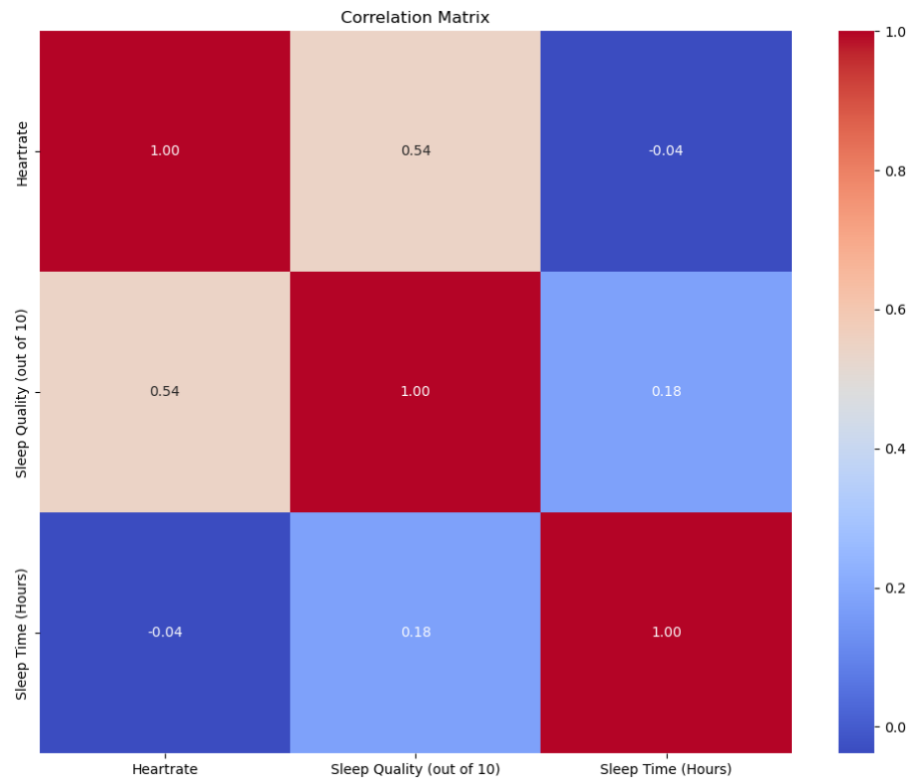


Figure 1. Correlation matrix between heart rate, sleep quality, and sleep time.

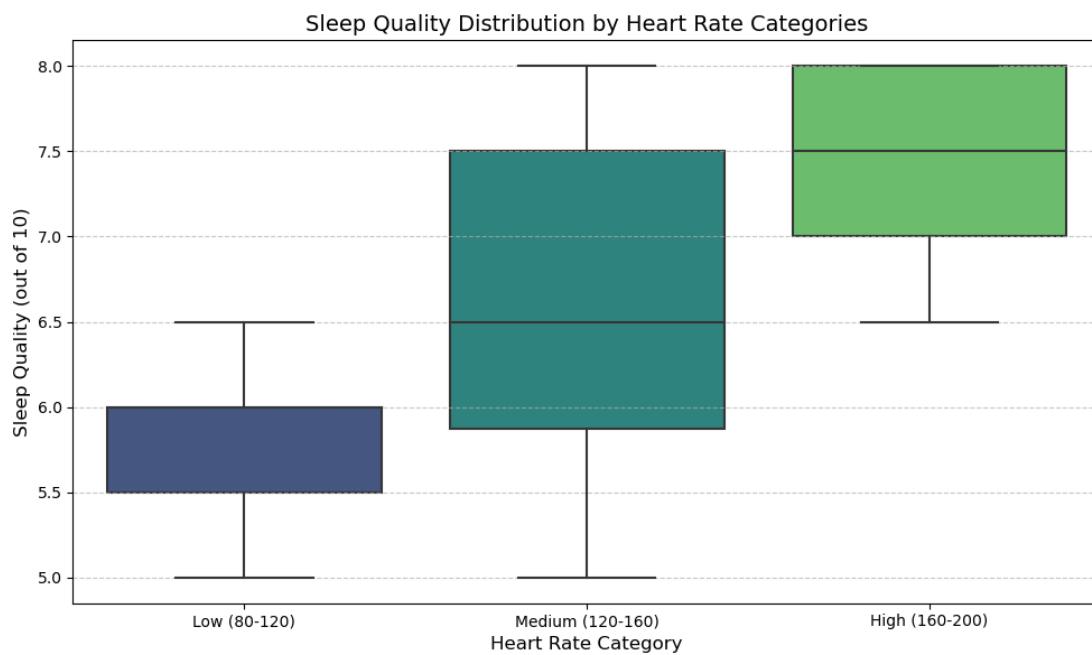


Figure 2. Sleep quality distribution over different heart rate categories.

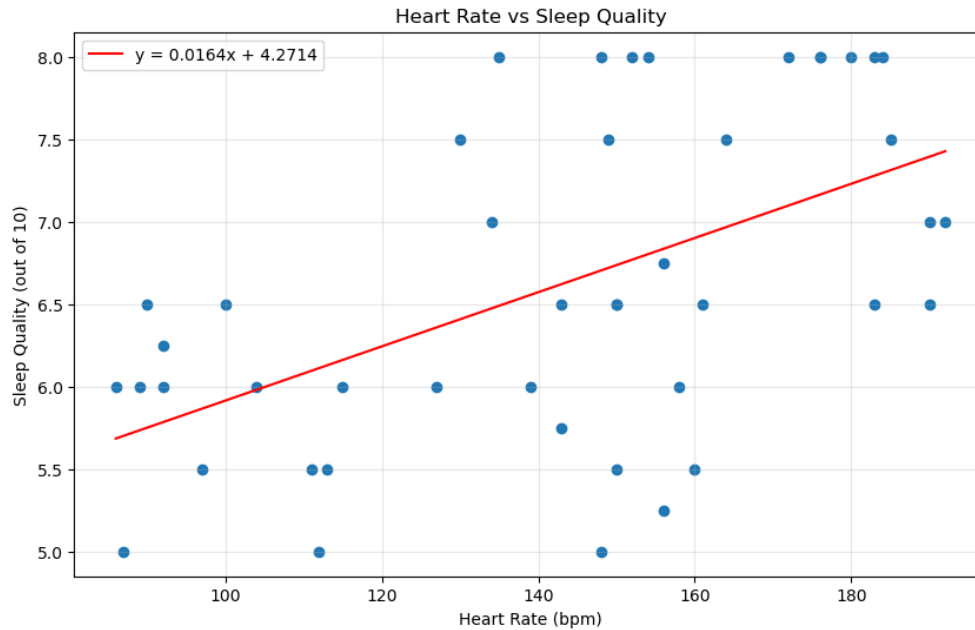


Figure 3. Plot of heart rate vs sleep quality.

3 Part 3: Plans for Remaining Weeks

I will continue to collect data for the remaining weeks. However, I will implement the change that was mentioned in part 1 implemented in part 2. That is that I will only collect the intensity of the exercise and the sleep quality each day instead of the time of day the exercise occurs. This way I can measure one independent variable and see if that variable has any correlation with my sleep quality. I have also noticed that I should have some more sample data for days with a low heart rate range. So, I will do some more days with low heart rate to get equal amounts of data.

4 Resources used to achieve this goal

Canvas: Homework template

Python Libraries: NumPy, pandas, matplotlib, seaborn, scikit learn

Excel Spreadsheets

5 References