**Origins:**

The idea for my hypothesis started when I began facing sleep issues last year. To understand my sleep patterns better, I started using an Apple Watch to track my sleep. Since wearing it to bed was a bit disruptive, I ended up recording more frequently on nights when I thought I wouldn't sleep well. I wanted to see if there was any pattern in my sleep data. However, as I collected more information, I began to doubt whether my data was affected by survivorship bias. Essentially, I realized that my data mostly reflected nights when I expected poor sleep, potentially skewing the overall analysis. So, my hypothesis revolves around the possibility that survivorship bias might be influencing the patterns I observe in my sleep data. It's crucial to address this doubt to ensure the reliability of my conclusions. Henceforth it is my hypothesis.

**Null Hypothesis (H0):**

There is no significant relationship between the frequency of recording sleep and the duration of sleep; an increase in the frequency of recording does not correlate with a decrease in the number of hours of sleep obtained.

Collecting the data:

I have diligently documented data from May 2023 to the present, yet there are intermittent gaps due to occasional lapses in daily recording.

**Processing the data:**

I extracted the data from my Apple Watch and discovered it was in XML format. After considerable effort, I opted to convert it to a CSV file, aligning with the format encountered in CS210 homework assignments. A YouTube tutorial guided me to a [website](https://www.ericwolter.com/projects/apple-health-export/) that effectively transforms Apple Health data into a CSV file specifically tailored for sleep data. Recognizing its utility, I incorporated this tool into my project workflow.

Website: https://www.ericwolter.com/projects/apple-health-export/

Given that Apple records data in various types, with the iPhone capturing only the time I am away from the phone at night, and the Apple Watch monitoring both heart rate and movement, I have opted to exclusively utilize data from my Apple Watch. Within the diverse categories provided by the Apple Watch data, I have chosen to visualize it on a chart to enhance clarity and facilitate a better understanding.

In terms of frequency classification, I've designated months with more than 15 recorded days as "frequent" and those with 15 days or less as "less frequent." Following a thorough analysis, I computed the p-value, and the results, as indicated in my code, reveal no significant correlation between the designated frequencies and the observed data.

In the code you can find the relative graphs, used libraries and methods that are explained by comments.