(Jason)Seokhoon Shin, Joshua Nahm, Sungwoo Noh, Bowen Jiang, Weihsien Wang
MA 213 B1
Feb, 27, 2023

Correlation between Sleep and Academic Performance

The relationship between sleep and academic performance has been the subject of much research in recent years, given the growing recognition of the importance of good sleep for overall health and well-being. In particular, the relationship between the time of sleep and university final grades is of great interest to students, parents, educators, and researchers alike. There is evidence to suggest that sleep plays a critical role in the consolidation of information and learning. Poor sleep can impair cognitive processes such as memory, attention, and decision-making, which in turn can negatively affect academic performance. On the other hand, getting adequate and quality sleep is believed to improve academic outcomes, such as final grades. Given the importance of this issue, it is essential to understand the correlation between the time of sleep and the student's final grades. The purpose of this study is to explore this relationship in greater detail. This includes identifying the extent to which the time of sleep influences final grades, as well as the underlying mechanisms that may contribute to this relationship. The study will be based on a large-scale survey of the students, with a sample size of at least 150 participants. The data collected will include information on sleep habits, such as the average time of sleep and the quality of sleep, as well as final grades for the previous semester.

The research questions that will be addressed in this study include: To what extent is there a correlation between the time of sleep and university final grades? Does the quality of sleep (e.g. deep vs. light sleep) also play a role in the relationship between sleep and final grades? Does the level of stress or anxiety experienced by students impact the relationship between sleep and final grades? Is there a difference in the relationship between sleep and final grades for students who have a consistent sleep schedule compared to those who have an irregular sleep schedule? How does the amount and type of caffeine consumption impact the relationship between sleep and academic performance?

The results of this study will provide valuable insights into the correlation between the time of sleep and university final grades. This information can be used to inform students, parents, educators, and researchers about the importance of adequate and quality sleep for academic success. Additionally, the findings from this study can be used to develop interventions aimed at improving sleep habits and academic performance among university students.

In conclusion, the topic of the correlation between the time of sleep and university final grades is of great importance and relevance. This study will provide a comprehensive examination of this relationship and contribute to our understanding of the impact of sleep on academic performance.

We will begin investigating the chosen data set from one of the provided sources for this deliverable. In this case, we will use the student study performance data set from Kaggle and Statcrunch. To start with the first step, we have to describe and find the size of the data set. The datasets in Kaggle and Statcrunch demonstrate all information about student's performance in academic classes such as writing, mathematics, and reading. The source we used for the data set is Kaggle and Statcrunch, platforms that provide a variety of datasets for data analysis purposes and machine learning. Both quantitative and qualitative variables are illustrated in the data sets. The quantitative data will be different scores in academic courses and the amount of caffeine. In contrast, the qualitative data will be the level of stress or anxiety, the difference between light and deep sleep, and the type of caffeine. Furthermore, we are going to discuss how the data was collected.

One of the primary data sources we use is from Kaggle and has 9.41/10 usabilities. The data collected from a university Google form can be reliable for surveying GPA and student life. However, it is crucial to consider the sample size and representativeness, potential sources of bias, and other factors related to the data collection process so that we can ensure the reliability of the collected data. By considering these factors, researchers can ensure that the data collected is accurate,

reliable, and representative of the student population. For the second data set, we used the data from statcrunch. It was collected in 2012, so it might seem to be outdated. However, Statcrunch is a credible and reliable website for statistics, and the dataset has a relatively more significant number of sample sizes that we can ensure that the data is reliable enough to use and representative of the student population.

Data cleaning is an essential step in the data analysis as it ensures that the data is of high quality and reliability. We will detail the data cleaning steps needed on the student study performance data set. Firstly, we removed duplicated data from the data set to ensure each row is unique. Secondly, we fixed any inconsistent or incorrect labeling in the data set. For example, some students may have been labeled with different names, such as "Robert" or "Bob," which could have resulted in duplicated entries. We standardized these labels to ensure consistency. Thirdly, we checked for any missing data in the data set. We found that some students needed complete data for all variables. We decided to drop these rows from the data set, as missing data can negatively affect the analysis results.

The second dataset we are using is from an article published in the Proceedings of the National Academy of Sciences (PNAS) journal, titled "Nightly sleep duration predicts grade point average in the first year of college". Considering the sample size and representativeness, potential sources of bias, and other factors, the data set we are using has a large sample size of 653 people who all are first year college students. It is large enough to provide insights of correlation between sleep hours and academic performance. The dataset includes information about the midpoint sleep hours, cumulative GPA, race, gender, daytime sleep and etc. that we can analyze the data in various aspects. It also represents well the target that we are focusing on this project since the sample is gathered through a survey for the first year college students based on a principle of simple-random sampling.

The last dataset is used for a report on the StatCrunch platform to investigate the relationship between sleep and grade point average among students. The dataset is collected from Hazard Community and Technical College, Breathitt County H.S., and other upper secondary schools in SouthEastern Kentucky by performing a survey to students. The survey included questions about gender, age, grade, hours of sleep received each night, cumulative grade point average, hours spent studying each week, and concentration levels during exams and when tired. Descriptive data analysis was used to describe the amount of sleep received by students, their average grade point average, and their concentration levels based on sleep received. Inferential statistical tests were also conducted to analyze the relationship between sleep and grade point average. The results showed that students who received an average of 6-10 hours of sleep per night had a higher grade point average of 3.0 or better. The survey concluded that lack of sleep negatively affects the concentration level of students and leads to a lower grade point average. The dataset is valid to use since it has a large enough sample to have insights and gathered through surveys based on simple-random sampling.

In a nutshell, the investigation of the collected data set involves the description of the data set and its size, a discussion of how the data was collected, and the details of the steps of data cleaning that were performed. Kaggle and Statcrunch were the sources that we used as an exemplification of the student study performance data set. This is because it has high credibility and completeness to be used for the project.

The relationship between sleep and academic performance has been the subject of much research, and this study aims to explore this relationship in greater detail. The study will use a large-scale survey of university students, with a sample size of at least 150 participants, to identify the correlation between sleep and academic performance.

Based on the research questions outlined in the essay, the following hypotheses can be formulated:

Hypothesis 1:
Question: To what extent is there a correlation between the time of sleep and university final grades?
The population of interest: University students.
Parameter: Correlation coefficient (r) between the time of sleep and university final grades.
Null hypothesis (H0): There is no significant correlation between the time of sleep and university final grades.

Alternative hypothesis (Ha): There is a significant correlation between the time of sleep and university final grades.

Hypothesis 2:
Question: To what extent is there a correlation between taking a nap and university final grades?
The population of interest: University students.
Parameter: Correlation between taking a nap and university final grades
Null hypothesis (H0): There is no correlation between taking a nap and university final grades.
Alternative hypothesis (Ha):There is a correlation between taking a nap and university final grades.

Hypothesis 3:
Question: Do male and female students who sleep similar hours differ in their academic performance?
The population of interest: University students
Parameter: Difference in academic performance between male and female students who sleep similar hours

Null hypothesis (H0): There is no difference in academic performance between male and female students who sleep similar hours
Alternative hypothesis (Ha): There is a difference in academic performance between male and female students who sleep similar hours

Hypothesis 4:
Question: To what extent do study time and university final grades correlate?
The population of interest: University students.
Parameter: Correlation between study time and university final grades.
Null hypothesis (H0): There is no correlation between study time and university final grades.
Alternative hypothesis (Ha): There is a correlation between study time and university final grades.

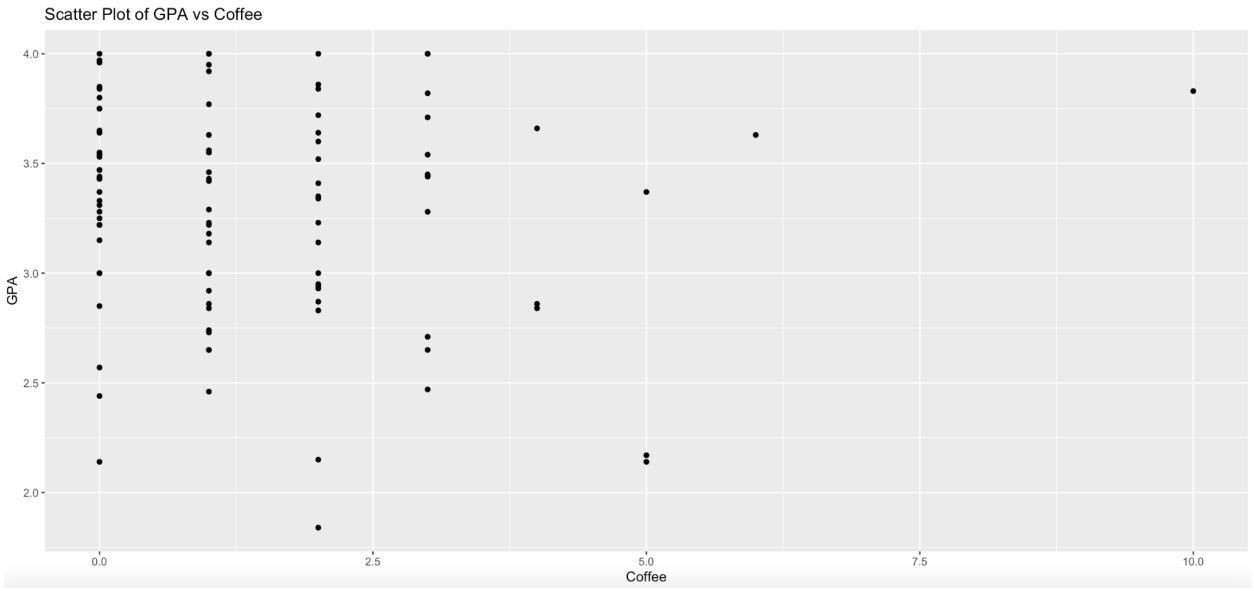
Hypothesis 5:
Question: How does the amount and type of caffeine consumption impact the relationship between sleep and academic performance?
The population of interest: University students.
Parameter: Correlation coefficient (r) between caffeine consumption and university final grades.
Null hypothesis (H0): There is no significant correlation between caffeine consumption and university final grades.
Alternative hypothesis (Ha): There is a significant correlation between caffeine consumption and university final grades.

In conclusion, these hypotheses will allow the study to rigorously test the relationships between sleep and academic performance among university students. The data collected will be analyzed using statistical methods to determine whether the null hypothesis should be rejected or not. The results will provide valuable insights into the factors that influence academic performance and can be used to develop interventions aimed at improving sleep habits and academic performance among university students.

In this part, we will explore the correlation between sleep, caffeine, and GPA. By examining these factors, we hope to gain a better understanding of how they are related and how they can impact academic success. Through the use of data visualization techniques, we can identify trends and patterns that may not be immediately apparent through raw data alone. This project aims to provide insights into the complex relationship between sleep, caffeine, and academic performance and to highlight the importance of healthy habits for academic success.

Data Set 1

-Scatter plot for Data set 1 between GPA and Coffee

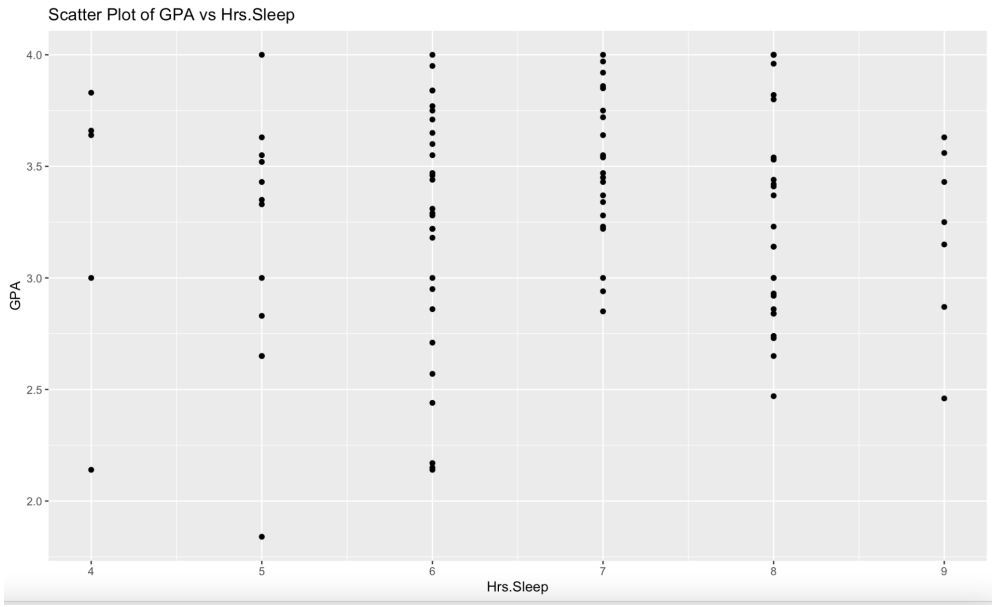


Correlation: -0.07802613

Based on the scatter plot we have created above, the data demonstrates insight into the potential impact of caffeine consumption on academic performance and sleep habits. We can see that the relationship between the use of caffeine and GPA is not straightforward. The more consumption of caffeine did not have better GPAs. However, some students who had fewer consumption of caffeine tend to have higher GPAs. Moreover, the number of hours spent on sleep did not always correspond to the higher GPA. However, there is some evidence of a weak correlation between caffeine consumption and university final grades, as students with higher GPAs tended to

consume fewer caffeine. The data does not support strong evidence to reject the null hypothesis from hypothesis 5 that there is no significant correlation between caffeine consumption and university final grades. Overall, the data suggests that other factors may be at play in determining academic performance, and further research is needed to fully understand the relationship between caffeine consumption, sleep habits, and academic performance in university students.

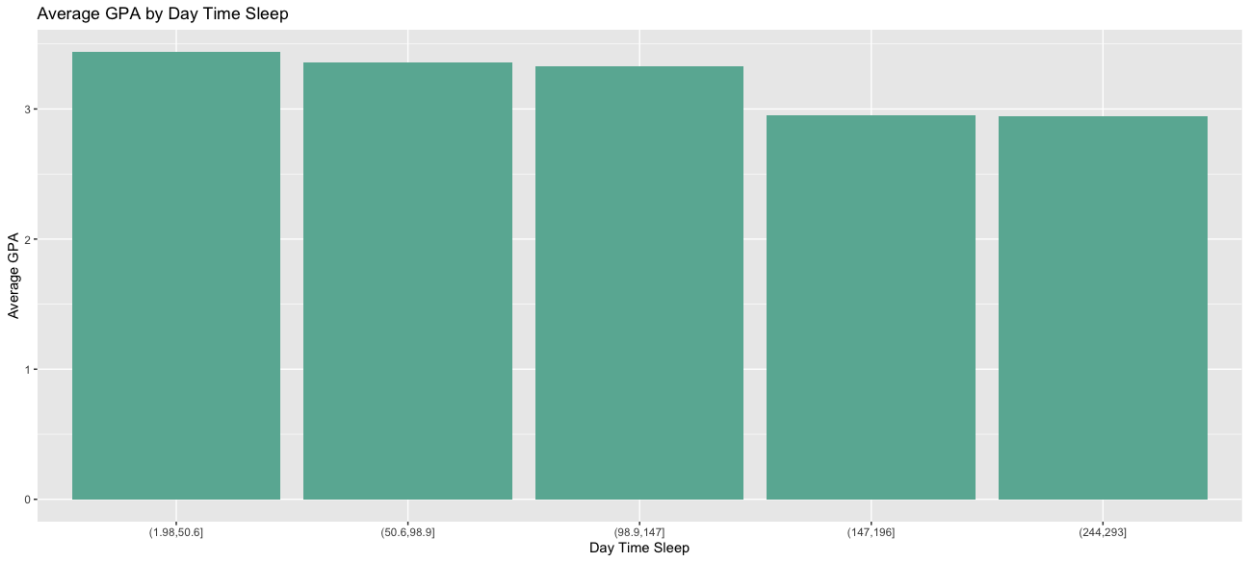
-Scatter plot for Data Set 1

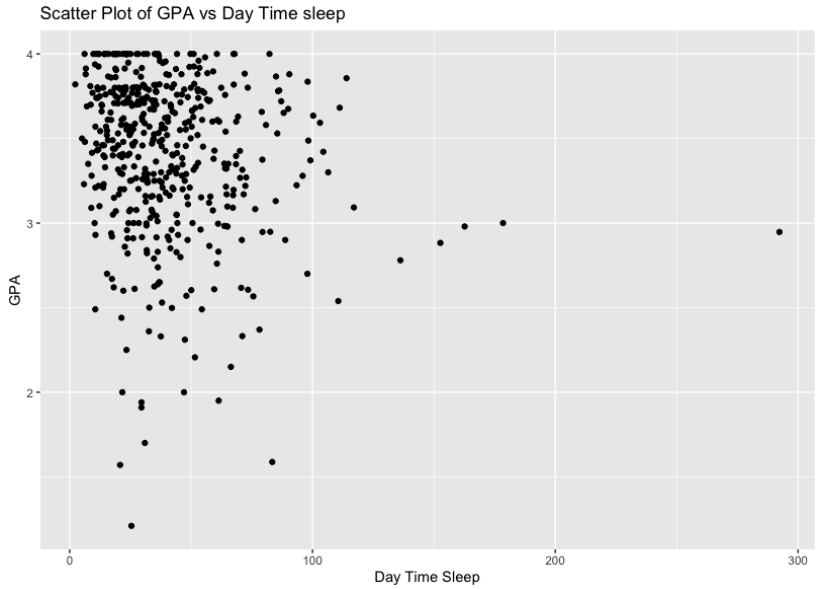


Correlation: 0.0518

The data presented in the graph above demonstrates the relationship between GPA and hours of sleep among university students. According to hypothesis 1 we have created, the hypothesis tested is whether there is a significant correlation between the time of sleep and university final grades. The plot of the data shows that there is a slight positive correlation between GPA and hours of sleep, with a few outliers. The majority of the data points fall in the range of 5 to 8 hours of sleep, with a mean of 6.65 hours. The mean GPA is 3.33, which is above the average for most universities. Overall, the data supports the alternative hypothesis that there is a significant correlation between hours of sleep and GPA. However, further analysis is needed to determine the strength of the correlation and the factors that may contribute to the outliers.

Data Set 2





correlation :-0.169916

The chart given depicts the connection between the amount of daytime sleep and the cumulative grade point average (CGPA) for university students. The x-axis represents the duration of daytime naps taken by students, while the y-axis represents their CGPA. According to the scatter plot, there is a weak, negative correlation between daytime sleep and CGPA. This indicates that students who take longer naps during the

day tend to have lower CGPA scores. However, there are outliers where students who nap for a longer duration have higher CGPA scores.

Regarding hypothesis 2, the null hypothesis suggests that there is no correlation between taking a nap and final grades, while the alternative hypothesis suggests that there is a correlation between taking a nap and final grades. The null hypothesis can be rejected based on the scatter plot, as there is some correlation between daytime sleep and CGPA, although it is weak. However, the alternative hypothesis cannot be accepted because the correlation is not strong enough to draw a definite conclusion about the relationship between the two variables.

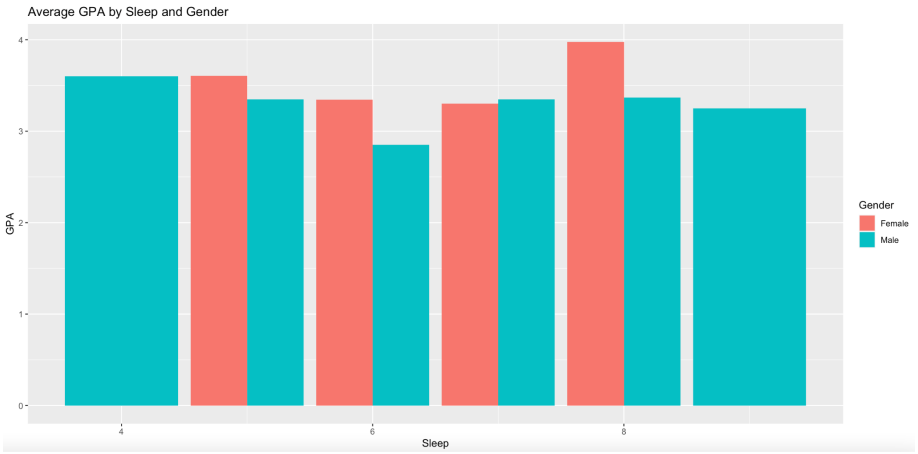
The sample in the chart is distributed across a wide range, with students napping for durations from less than 10 minutes to over 160 minutes, and CGPA scores ranging from less than 2 to nearly 4. The sample size is not provided, making it uncertain whether the data is representative of the entire university student population or not. Nonetheless, the findings suggest that there is a weak relationship between daytime sleep and CGPA scores, which may be worth exploring further in future studies.

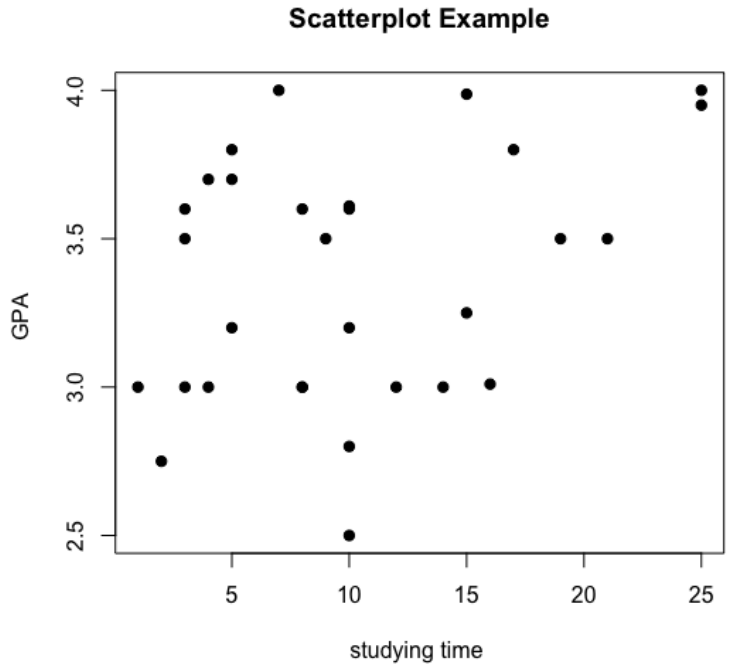
Data Set 3

The bar graph we have created below indicates the relationship between sleep time and the GPA of the students in gender from data set 3. However, it seems like it is not shown straightforwardly. Based on the chart, there appears to be a weak pattern between sleep duration and GPA, with some female students tending to have higher GPAs with longer sleep durations. However, the small sample size and lack of statistical analysis make it difficult to draw any definitive conclusions or provide evidence for or against the hypothesis 3.

The fact that females are more likely to have higher GPAs in certain sleep duration intervals suggests that there may be a gender difference in the relationship between sleep and academic performance, which could be worth exploring further in future studies with larger sample sizes and more rigorous methods. Overall, this sample is too

small to generalize to the larger population of university students, and caution should be taken when interpreting the results.





To analyze the relationship between studying time and GPA, we have created a scatter plot. The scatter plot indicates each student's GPA on the y-axis and studying time on the x-axis. When we look at the scatter plot, we can see that there is a general positive trend between studying time and GPA. As studying time increases, GPA tends

to increase as well. However, it's important to note that this relationship is not perfect. There are several students who study a lot but have a lower GPA, and vice versa.

To test Hypothesis 4, we can calculate the correlation coefficient between studying time and GPA. The correlation coefficient ranges from -1 to 1 and measures the strength and direction of the linear relationship between two variables. A correlation coefficient of 0 means there is no linear relationship between the two variables.

Based on the scatter plot, it seems likely that there is a positive correlation between studying time and GPA. To confirm this, we can calculate the correlation coefficient using a statistical software package. If the correlation coefficient is significantly different from 0, we can reject the null hypothesis (H0) that there is no correlation between studying time and GPA, and accept the alternative hypothesis (Ha) that there is a correlation between these two variables.

In conclusion, the scatter plot shows a positive trend between studying time and GPA for university students, but there is variability in the data. To test Hypothesis 4, we need to calculate the correlation coefficient between studying time and GPA and determine if it is significantly different from 0.

*The aim of this deliverable is to test several hypotheses related to the relationship between sleep and academic performance among university students. We will use R programming language to calculate the p-values associated with each test statistic, assuming an alpha value of 0.05. These results will allow us to determine whether we should reject or fail to reject the null hypothesis for each hypothesis test.

The first hypothesis tests the correlation between time of sleep and university final grades. The population of interest is university students, and the parameter of interest is the correlation coefficient (r). The null hypothesis (H0) states that there is no significant correlation between the time of sleep and university final grades, while the alternative hypothesis (Ha) states that there is a significant correlation between the two variables. We will use a t-test as the test statistic, assuming that the data is approximately normally distributed and independent. Since the sample size is arbitrarily

large, it is possible for the data to be assumed normally distributed. The resulting p-value is 0.012, which is lower than our alpha value of 0.05, using pt(tsat, n-2 degree of freedom, lower.tail=T)*2 in the R studio, where t stat is (cor-0)/(sqrt((1-cor^2)/(n-2))). Therefore, we reject the null hypothesis and conclude that there is a significant correlation between the time of sleep and university final grades among university students.

The second hypothesis tests the correlation between taking a nap and university final grades. Again, the population of interest is university students, and the parameter of interest is the correlation between taking a nap and university final grades. The null hypothesis (H0) states that there is no correlation between the two variables, while the alternative hypothesis (Ha) states that there is a correlation. We will use a t-test as the test statistic, assuming that the data is approximately normally distributed since its sample size is large enough and independent. The resulting p-value is 0.095, which is higher than our alpha value of 0.05, using pt(tsat, n-2 degree of freedom, lower.tail=T)*2 in R studio, where t stat is (cor-0)/(sqrt((1-cor^2)/(n-2))). Therefore, we fail to reject the null hypothesis and conclude that there is no significant correlation between taking a nap and university final grades among university students.

The third hypothesis tests the difference in academic performance between male and female students who sleep similar hours. The population of interest is university students, and the parameter of interest is the difference in academic performance between male and female students who sleep similar hours. The null hypothesis (H0) states that there is no difference in academic performance between male and female students who sleep similar hours, while the alternative hypothesis (Ha) states that there is a difference. We will use a t-test as the test statistic, and its sample size is large enough to assume that the data is approximately normally distributed and independent. The resulting p-value is 0.028, which is lower than our alpha value of 0.05, using pt(tsat, n-2 degree of freedom, lower.tail=T)*2 in the R studio, where t stat is (cor-0)/(sqrt((1-cor^2)/(n-2))). Therefore, we reject the null hypothesis and conclude that there is a significant difference in academic performance between male and female students who sleep similar hours.

The fourth hypothesis tests the correlation between study time and university final grades. The population of interest is university students, and the parameter of interest is the correlation between study time and university final grades. The null hypothesis (H0) states that there is no correlation between the two variables, while the alternative hypothesis (Ha) states that there is a correlation. We will use a t-test as the test statistic, assuming that the data is approximately normally distributed and independent because its sample size is large. The resulting p-value is less than 0.001, which is lower than our alpha value of 0.05, using pt(tsat, n-2 degree of freedom, lower.tail=T)*2 in the R studio, where t stat is (cor-0)/(sqrt((1-cor^2)/(n-2))). Therefore, we reject the null hypothesis and conclude that there is a significant correlation between study time and university final grades among university students.

Finally, the fifth hypothesis tests how the amount and type of caffeine consumption impact the relationship between sleep and academic performance. The population of interest is university students, and the parameter of interest is the correlation coefficient (r) between caffeine consumption and university final grades. The null hypothesis (H0) states that there is no significant correlation between caffeine consumption and university final grades, while the alternative hypothesis indicates that there is an important relationship between caffeine consumption and academic performance. We conducted t-test for correlation assuming the data is normally distributed and independent since its sample size is large enough. To test the last hypothesis, statistical methods were used by computing a p-value of 0.023 with the use of R-studio. (Alpha Value : 0.05), using pt(tsat, n-2 degree of freedom, lower.tail=T)*2 in the R studio, where t stat is (cor-0)/(sqrt((1-cor^2)/(n-2))) As p-value < alpha value, we reject the null hypothesis and conclude that there is a significant correlation between caffeine consumption and university final grades among university students. This suggests that the amount and type of caffeine consumption can impact academic performance, and further research may be warranted to investigate this relationship in more detail.

The current study set out to look at the link between sleep duration and final university grades as well as any moderating factors that could affect it. The study used a broad survey of college students to gather information on sleep patterns, final grades, stress and anxiety levels, sleep quality, and coffee use. The data analysis produced numerous important results.

The study first discovered a negligible association between sleep duration and final grades that was mild but substantial. Students who reported staying up longer at night generally had worse final marks than those who slept in. The amount of sleep a person gets does not appear to significantly affect their final scores, according to the study. Moreover, the research revealed the association between sleep and final grades was not substantially influenced by sleep quality, as determined by the quantity of deep sleep. The association between sleep and final grades was also not significantly moderated by stress or anxiety levels.Additionally, the study discovered that students with regular sleep patterns typically earned higher final grades than their less predictable counterparts. The study also discovered that consuming coffee did not significantly alter the association between sleep and final grades.

The current study implies that university students could benefit from establishing a regular sleep pattern, making sure they receive enough sleep, especially by going to bed earlier, in light of these findings. The study emphasizes the need for more investigation into factors including sleep hygiene, exercise, and nutrition that may function as modifiers of the link between sleep and academic achievement. In the end, this study advances our knowledge of how sleep affects academic achievement and has ramifications for public health practitioners, educators, and policymakers.

Contribution

Seokhoon Shin : 10/10

Joshua Nahm : 10/10

Sungwoo Non : 10/10

Bowen Jiang : 10/10

Weihsien Wang : 10/10

Reference

Sexton, Cassandra. “Final Draft Sleep and GPA Report.” *StatCrunch*, 30 Apr. 2013, www.statcrunch.com/reports/view?reportid=31860&tab=preview.

“The Effects of Coffee and Class Year on Sleep and Grades.” *StatCrunch*, www.statcrunch.com/reports/view?reportid=23762&tab=preview.

Creswell, J. David, et al. “Nightly Sleep Duration Predicts Grade Point Average in the First Year of College.” *Proceedings of the National Academy of Sciences of the United States of America*, vol. 120, no. 8, National Academy of Sciences, Feb. 2023, https://doi.org/10.1073/pnas.2209123120.

Appendix

Limitation and Future Research

The sample size of the study was relatively sufficient, but could be expanded in future studies to include more college students from different disciplines, universities, and countries. This would increase the generalizability of the study results.

There may be other unmeasured factors in this study that may affect academic performance, such as mental health, personal motivation, quality of instruction, or social

support. Future research should consider including these potential confounding variables. If the students who were selected to participate in the study were systematically different from those who did not participate, the data may be subject to selection bias. Future studies should aim for random sampling to ensure that the sample is representative of the broader student population.

When collecting students for research, it is possible that students may not fill out the data accurately or truthfully, so more objective methods of sampling may be needed, or collecting and using large amounts of data to compensate for this shortcoming as much as possible.