Sleep Effect on GPA in University Students

Seyoung Lee

October 13, 2023

1. Introduction

Adequate sleep is crucial for a student's overall academic performance in college, and with a lack of sleep potentially impacting a student's GPA, it is essential to understand to what degree of impact sleep has on a student's GPA. To measure if lack of sleep influences a student's grades, a total of 634 students from Carnegie Mellon and two other universities were recruited to track the average hours of sleep each participant slept each night for a month during their spring semester. At the end of the semester, researchers obtained each participant's semester grades to explore the relationship between sleep and grades in university students.

More specifically, the following three research questions were explored:

1. What is the association between sleep time and GPA? Is there evidence that students who sleep less indeed get lower GPAs?
2. The Vice Provost of Carnegie Mellon expects that students would sleep 2 hours less, on average, if they have to get up for an 8am class. What average GPA effect do we expect to see with 2 hours less sleep?
3. Based on the data, can we conclude that the GPA change is caused by less sleep, or could their be other reasons for the relationship?

Because participants' schedules remained private, it is difficult to assume that the implementation of early classes directly impacts grades. But based on past research on sleep time with students, if students with 8 am classes were to sleep 2 hours less, on average, we theorize that early-morning classes set with less sleep time does indeed have an impact on student grades.

In our analysis, we find that there is indeed a slight association between GPA and sleep time. However, we also find that GPA from previous semesters is the better variable for predicting current GPA. Additionally, two hours less sleep is associated with a slight decrease in GPA in students, but it cannot be concluded that GPA change is caused by less sleep due to a multiple of other confounding variables unaccounted for.

2. Exploratory Data Analysis & Data Summary

2a. Overall Data Information

There were three main variables collected for each n=634 participant pool:

TotalSleepTime: The average time the student spent asleep each night, not counting naps, in minutes.

term_gpa: The student's GPA (out of 4.0) for the classes they took in the semester being studied.

cum_gpa: The student's GPA (out of 4.0) for the semesters before the one being studied (or first-year students' fall semester grades).

(citation: Reinhart, Alex, "Data Exam 1", 2023.)

2b. Exploring Total Hours of Sleep (TotalSleepTime.Hours)

For purposes of readability, we converted the original variable for average sleep time in minutes to total sleep time in hours, or TotalSleepTime.Hours. Below is the distribution of the average sleep time in hours each night.

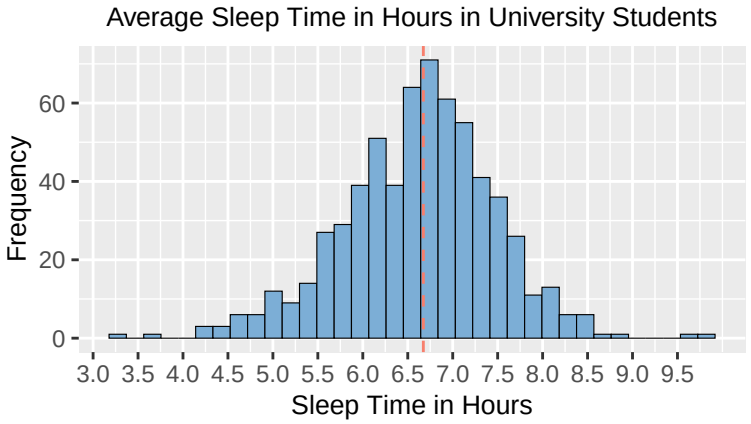


Figure 1.

The distribution shown in Figure 1 displays the average sleep time in hours; we see that the distribution is unimodal and normally distributed. Based on this variable, students received an average of 6.62 hours and a median of 6.67 hours of sleep. The middle 50 percent of students received 6.12 to 7.17 hours of sleep. As the median line shows, we also see that the distribution is relatively symmetrical.

2c. Exploring Student GPAs (term_gpa, cum_gpa)

There were two main variables to track grades in the participants: the students' current semester GPA (term_gpa) and the students' GPA in the previous semester (cum_gpa).

With these two variables, we would be able to measure if there are significant differences in previous and current GPAs in the same student, which may be a result of confounding variables such as schedule changes or motivation.

Figure 2 shows the distribution of GPAs for both the current semester and previous semester.

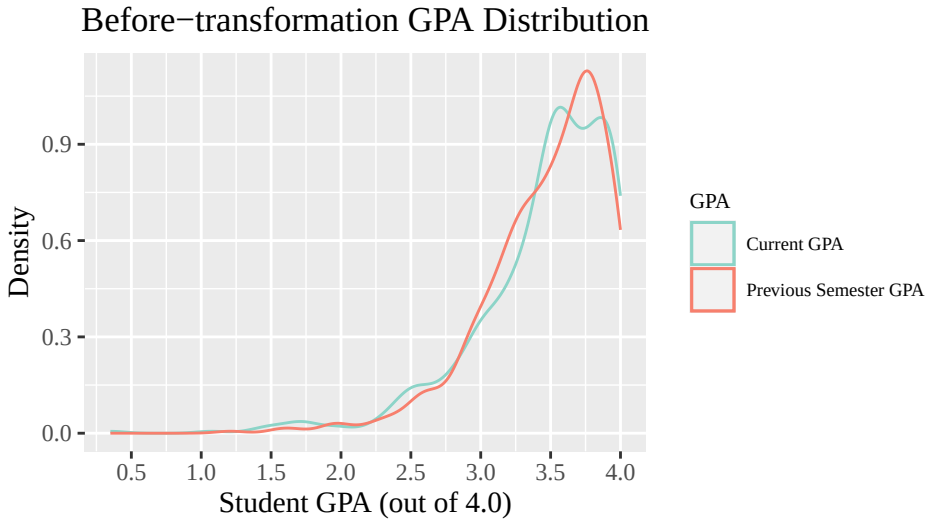


Figure 2.

As Figure 2. shows, student GPA is hightly left skewed, such that it does not appear to have a normal distribution. In fact, it is strongly negatively skewed with high mean and median values. To further test for

normality, we calculate the skewness of each GPA dataset: The current semester GPA (term_gpa) yields a -1.73 skewness, and the previous semester GPA (cum_gpa) yields a -1.35 skewness. This indicates a heavily negatively (left) skewed dataset, as a skewness closer to zero conveys symmetry.

In order to create normality in our data, we apply an inverse transformation for negatively skewed data. If we apply the same skewness calculation on each GPA dataset (Figure 3.), we see that the skewness has significantly increased, such that it is closer to zero. The transformed current semester GPA (term_gpa) now yields a -0.08 skewness, and the transformed previous semester GPA (cum_gpa) now yields a -0.002 skewness.

(citation: https://www.datanovia.com/en/lessons/transform-data-to-normal-distribution-in-r/)

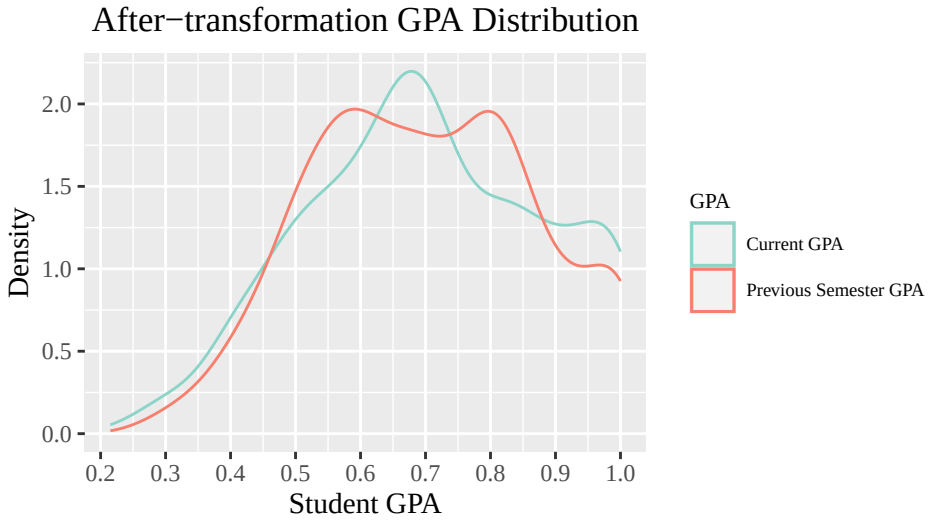
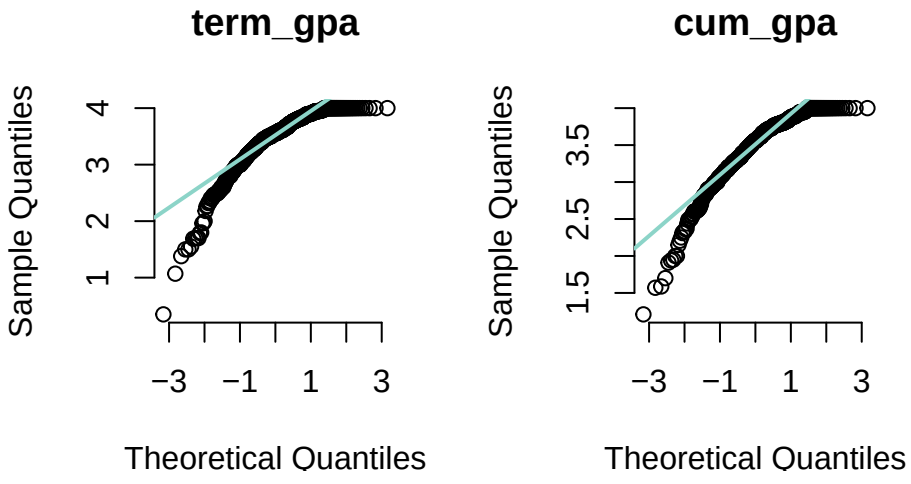
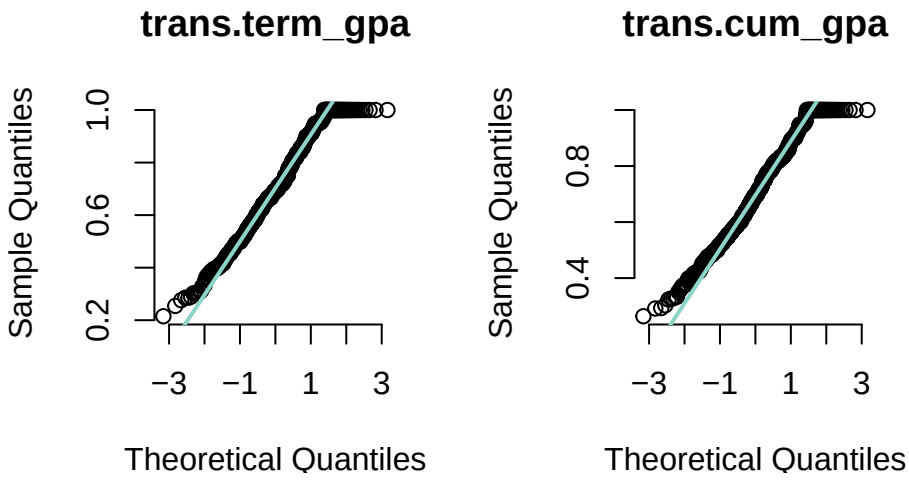


Figure 3.

After the transformation, although still slightly skewed left, the dataset seems much more normal, with a mean and median of about 0.70 in both the current semester GPA (term_data) previous semester GPA (cum_data).

We can further test for normality through Normal QQ plots. The first row shows the original GPA data, while the second row shows the transformed GPA data. The transformation has positively impacted the normality of the data.





With our transformed data, we now perform an initial bivariate EDA to visually test if a relationship between GPA and hours slept in students. We plot the effect of hours of sleep in a simple linear regression model, and we see in the figure below that GPA by Total Hours Slept suggests some linearity. This further demonstrates that the transformation was successful, such that there may be some association between hours slept and GPA.

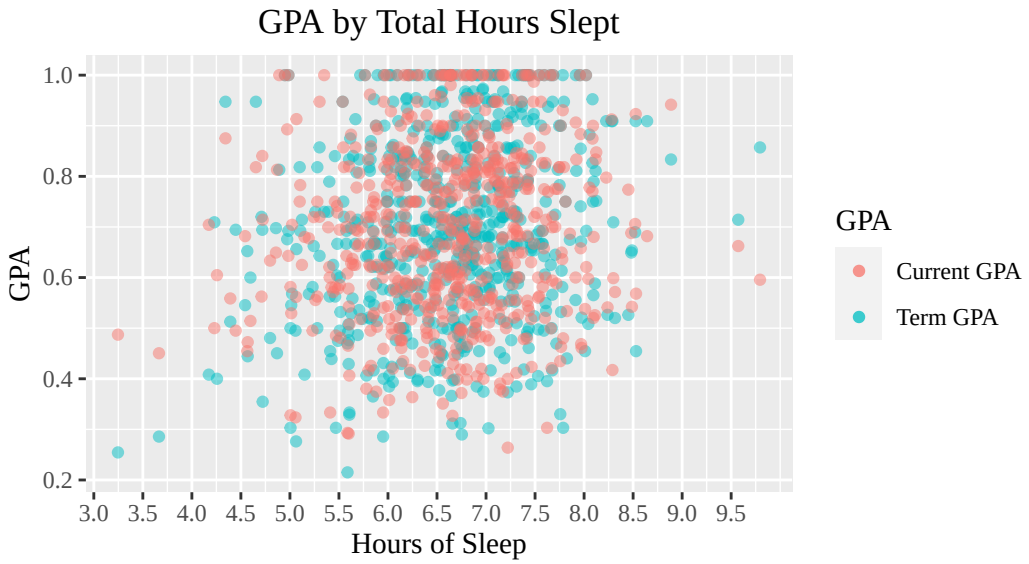


Figure 4.

2d. Limitations

We also recognize that the data is limited in several ways. First, applying transformations, especially with heavily skewed dataset, will affect its interpretations later on since significance in data is now based on the inverse transformation, rather than the original data itself. Additionally, the population of the dataset favors university students from three universities, which may have similar to varying degrees of academic culture unique to each university. Hence, although non-normality may be bypassed by the Central Limit Theorem (n>30), it is still important to note that the distribution of the GPAs may be favorable to the recruited populations.

3. Methods

3a. Research Question 1 Methods

> *What is the association between sleep time and GPA? Is there evidence that students who sleep less indeed get lower GPAs?*

To test whether or not there is an association between sleep time and GPA, we take note of the linearity between hours of sleep (TotalSleepTime.Hours) and GPA (term_gpa and cum_gpa) suggested in Figure 4.

Which variables are most correlated with one another?

Below are the correlations between each of the three variables (term_gpa, cum_gpa, and TotalSleepTime.Hours). The largest correlation between variables was previous GPA and current GPA, while the correlation between GPA and sleep hours is relatively small.

Table 1: Figure 5. Correlations Between Variables

Variable	Correlation
Sleep Hours vs. Current GPA	0.1793194
Sleep Hours vs. Previous GPA	0.1109871
Previous GPA vs. Current GPA	0.6932214

What do the residuals in each model indicate about the significance of the model?

Homoskedasticity refers to the equal scatter of residuals at different levels of TotalSleepTime.Hours. We test homoskedasticity because it is an essential assumption on linear regression, such that the residuals are evenly distributed at different levels of the response variable.

Below are the plots for the residuals yielded for two linear models: (1) Hours slept vs. Current GPA and (2) Hours slept & Previous GPA vs. Current GPA.

While the figure on the left does not have homoskedastic residuals, the right figure (measuring both cumulative GPA and sleep hours as predictor variables) appears to have a reliable variance throughout different levels of the response variable.

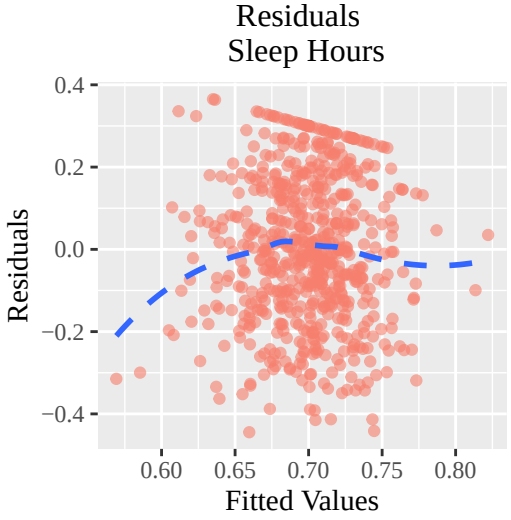


Figure 6.

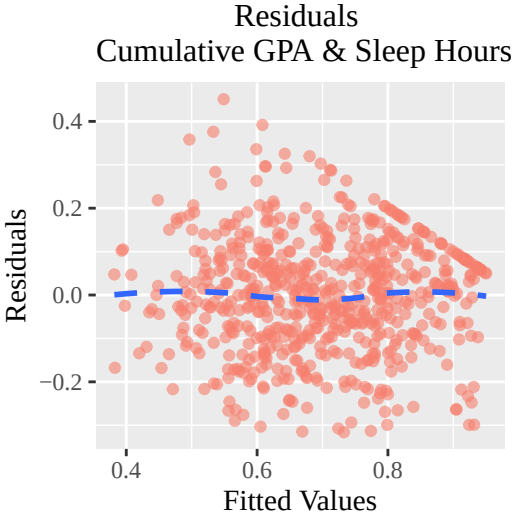


Figure 7.

Which linear model best predicts current GPA (term_gpa)?

The correlational data suggests that previous GPA is most correlated with current GPA (term_gpa) over sleep time hours (TotalSleepTime.Hours). Additionally, if sleep time remains the only predictor variable, we see in the residual plots, that it does not pass homoskedasticity while the other model (with both cumulative GPA and sleep time as predictors) achieved such tests.

> The result section will include the regression model with two predictors (TotalSleepTime.Hours & cum_gpa) and one response variable (term_gpa). With the three given variables, this model will most likely be able to explain if there is an association between sleep time and GPA (4a. Resarch Question 1 Results).

Confidence Intervals, Significance Testing, and Standardized Model

We first performed confidence intervals to measure the uncertainty in the association between sleep time in hours in our model. This will determine the precision of the parameter estimate, as a confidence interval suggesting more uncertainty indicates an unsatisfactory model.

After, we performed significance testing on all the parameters of the model. This indicates whether or not the predictor variables has a statistically significant effect on the response, in this case current GPA. Because this significance testing was based on both predictor variables, we standardized the model, such that comparison between the estimates of the predictor variables is reliable.

3b. Research Question 2 Methods

> *What average GPA effect do we expect to see with 2 hours less sleep?*

In order estimate the average GPA effect, we first performed a simple beta coefficient calculation, such that the estimate for total sleep time in hours is reduced by 2 hours less sleep respectively. Additionally, we perform a confidence interval in order to test the uncertainty of the value representing this effect. This will not only provide an estimate to how average GPA is affected by 2 hours less of sleep, but it will also provide a range of plausible values that can demonstrate the where the true parameter is likely to fall and outline the extent of uncertainty.

3c. Research Question 3 Methods

> *Based on the data, can we conclude that the GPA change is caused by less sleep, or could their be other reasons for the relationship?*

This question involves correlation vs. causation. Correlation does not imply causation, and because there is no form of experimental study design methods, we explain why we cannot assume GPA change is caused by less sleep relationship further in the results section.

4. Results

4a. Research Question 1 Results

> *What is the association between sleep time and GPA? Is there evidence that students who sleep less indeed get lower GPAs?*

To assess the association between sleep time and GPA, we look at the regression model with the two predictors (Sleep time in Hours & Cumulative GPA) below. We see that that the estimates of both predictor variables are positive.

Table 2: Figure 8. Linear Model Summary

Predictor	Estimate	Std_Error	T_Value	P_Value
Intercept	0.045105	0.044115	1.022	0.306969

Predictor	Estimate	Std_Error	T_Value	P_Value
Sleep Time in Hours	0.02232	0.00615	3.62775	0.000309019639926281
Cumulative GPA	0.72605	0.03043	23.85834	3.67631191796426e-90
Overall	*	*	*	< 2.2e-16

Additionally, the multiple linear regression model in Figure 9 below shows a clear positive linear relationship, which is represented by the following linear regression equation:

$$Y = 0.045105 + 0.02232X_1 + 0.72605X_2 \quad (1)$$

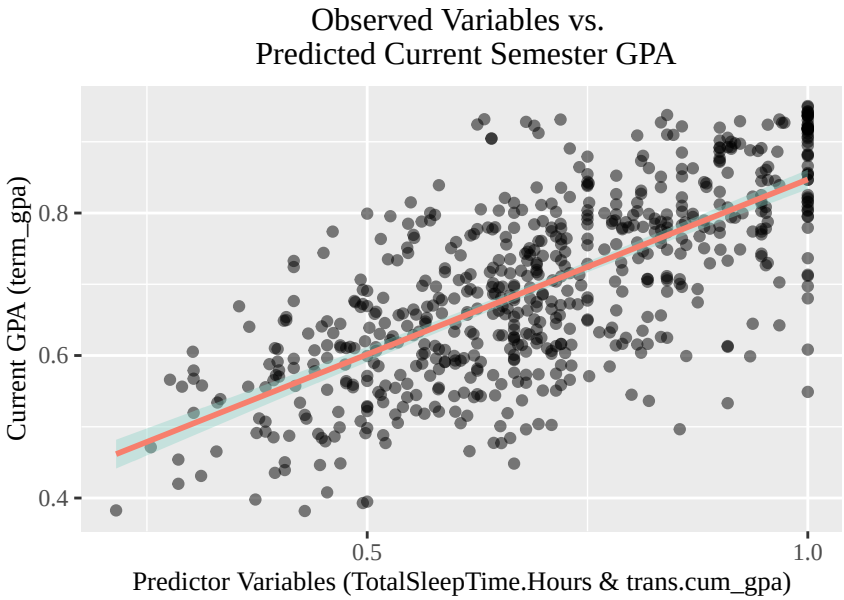


Figure 9.

How much uncertainty is there in the association between sleep time and term_gpa?

However, before finalizing a linear regression equation, it is important to note the uncertainty with confidence intervals and significance with hypothesis testing, as further explained in the methods section (3a).

Based on summary reports on the regression model as well as confidence interval predictions, we assume the following:

> Getting an additional hour of sleep is associated with having a current semester GPA that is $\hat{\beta}_1 = 0.022$ units higher, on average (95% CI[0.01,0.034]).

Are the predictors useful in predicting the current semester GPA?

Based on summary reports on the regression model, specifically the beta coefficients and p-values, we create a null and alternative hypothesis to check for significance:

> $H_0 : \beta_1, \beta_2 = 0$
>
> $H_1 : \beta_1$ or $\beta_2 \neq 0$
>
> There is strong evidence that at least one parameter, total sleep time in hours (TotalSleepTime.Hours) or cumulative GPA (cum_gpa), is associated with current GPA (term_gpa) (F=304.5, p<2.2e-16).

To find *which* predictor is better at predicting the students' current GPA, the initial parameter coefficients cannot be compared because they have different units and scales (GPA vs. Hours). Since we have established

if the predictors are useful for predicting the current semester GPA, we will calculate the standardized coefficients to compare the predictors themselves.

Table 3: Figure 10. Linear Model Summary Wtih Standardized Estimates

Predictor	Estimate	Std_Error	T_Value	P_Value
Sleep Time in Hours	0.02232	0.00615	3.62775	0.000309019639926281
Cumulative GPA	0.72605	0.03043	23.85834	3.67631191796426e-90
Overall	*	*	*	< 2.2e-16
Scaled Sleep Time Hours	0.1037	2.857e-02	3.628	0.000309
Scaled Cumulative GPA	0.6817	2.857e-02	23.858	< 2e-16

Figure 10 includes the standardized estimates for total sleep time in hours and cumulative GPA. Since it standardized, we can now compare the estimates:

> A one standard deviation increase in hour is associated with a 0.1037 increase in current GPA, assuming cumulative GPA is held constant.
>
> A one standard deviation increase in cumulative GPA is associated with a 0.6817 standard deviation increase in current GPA, assuming sleep time is held contant.

We see in our estimates that the cumulative GPA is more highly associated with current GPA over sleep time in hours. With this information, we are finally able to answer the research question.

What is the association between sleep time and GPA? Is there evidence that students who sleep less indeed get lower GPAs?

> Based on the confidence intervals, getting an additional hour of sleep is associated with having a current semester GPA that is $\hat{\beta}_1 = 0.022$ units higher, on average (95% CI[0.01,0.034]).
>
> Although there is some statistically significant association between sleep and GPA, as the linear regression model shows, cumulative GPA was a better indicator in predicting current GPA over hours of sleep a student received.

4b. Research Question 2 Results

> *What average GPA effect do we expect to see with 2 hours less sleep?*

Table 4: Figure 11.

	Effect
TotalSleepTime.Hours	-0.0446466

To estimate the average current GPA effect on 2 hours less sleep each night, we use the previous beta coefficient for total sleep time in hours (TotalSleepTime.Hours) in the linear regression model for predicting its effect on the current GPA. To test its uncertainty, a confidence interval is also shown below.

Table 5: Figure 12.

	Lower	Upper
TotalSleepTime.Hours	-0.0567303	-0.0325628

Based on the average effect size and confidence interval calculation, we assume the following:

On average, we expect that 2 hours less sleep each night is associated with a decrease in current GPA by -0.0446466 units, (95% CI[-0.0567303, -0.0325628]).

4c. Research Question 3 Results

Based on the data, can we conclude that the GPA change is caused by less sleep, or could their be other reasons for the relationship?

No, we cannot conclude that GPA change is *caused* by less sleep. There are a variety of other reasons that may explain this relationship, such as stress and mental health fluctuations, motivation adjustments throughout the semester, or student preferences in time of day to work. All of these factors can directly affect either hours of sleep or GPA.

Additionally, although there is evidence of correlation in GPA change and less sleep, a conclusion regarding causation cannot be made unless there is random assignment. In exploring the correlation analysis and standardized model, we also concluded that cumulative GPA is a better indicator for GPA change than hours of sleep. This might be because those who receive better grades previously do better in the next semester, which may mean that is more dependent on the student and their academic goals rather than hours of sleep students receive.

Hence, due to these confounding variables and lack of experimental data collection, we cannot assume that lack of sleep cases GPA change.

Conclusion

Overall, there does seem to be some correlation between GPA and sleep hours, but to answer the research questions, we state the following:

1. On average, an additional hour of sleep accounts for 0.022 units higher in current semester GPA. Although the correlation is significant and reliable, we found that the other variable (previous semester GPA) is better at predicting if students who get less sleep indeed get lower GPAs.
2. On average, 2 hours less sleep is associated with a decrease in GPA by -0.0446466 units. However, this does not indicate whether or not the implementation of 8 am classes is associated with less sleep.
3. We cannot conclude that GPA change is caused by less sleep due to various confounding variables not controlled for in an experimental study design.

The main limitations of our analysis comes from the abundance of potential confounding variables. The relationship between GPA and sleep hours is affected by a variety of different factors, such as sleep quality, sleep schedule (such as naps), physical activity, or mental health. All of these factors that can potentially affect either the predictor or response variables were unaccounted for. In future analysis, it may be important to look at the other variables in the sleep dataset, or within the collection process, also measure academic motivations.