36-401 Data Exam 1: Examining the Influence of Sleep on GPA

Anusha Bhat

October 13, 2023

1. Introduction

For this report, we are tasked by the Vice Provost of Carnegie Mellon University (CMU) to explore the effect of sleep on GPA. CMU has begun to offer classes starting at 8 am and the Vice Provost would like to know if these morning classes cause students to perform worse academically. Faculty and students at CMU theorize that those who have to attend these early classes get less sleep than usual, resulting in poor cognitive performance due to tiredness. Hence, we hypothesize that there may be a strong positive association between sleep and GPA.

A group of researchers conducted a study at CMU and two other universities where they enlisted 634 participating students to wear fit-bits to track their average sleep time each night for a month. The students continued their normal routines and there is no information on if a given student actually had 8 am classes or not. We also note that the study only had freshman participants. The researchers were also able to obtain the student's term and cumulative GPAs at the end of the semester the study took place in. It is important to investigate the effect of sleep on academic performance, especially GPA, since the goal of an institution is to provide the best environment for a student to achieve academic success. Getting a good night's sleep and plenty of rest is vital to maintaining good cognitive health which is necessary to understand course material. Hence, exploring the influence of sleep on GPA may help a university to create class schedules that optimize a student's sleep time and its impact on their GPA.

To explore the Vice Provost's broad question, we will address three specific research questions. Our first question will focus on the association between sleep and GPA and if students who sleep less have lower GPAs. In addition, the faculty and students at CMU believe that

early morning classes will cause a student to sleep two hours less on average, hence, we will explore what we can expect the impact of two hours of less sleep to be on average GPA for our second question. Finally, our third research question will address if there are potentially other explanations for the relationship we observe between sleep and GPA. Through our analysis, we find that sleep is positively associated with both term and cumulative GPA, however, we can not conclude causality without further research.

2. Exploratory Data Analysis

In our "study" data set, we have several variables such as student race, gender, and demographic, cumulative grade point average (GPA), term GPA, total number of units taken during the semester of the experiment, daytime total sleep duration, night time total sleep duration, and more that were collected for 634 students. For the purposes of this report, we will focus on the following three variables: cumulative GPA, term GPA, and total sleep time. Since we want to investigate if sleep is associated with GPA, we can use these variables to model any potential relationships with total sleep time as the predictor and term GPA and cumulative GPA as the responses respectively. Looking at other variables that are not related to sleep will not help us determine any potential relationships that can answer our research questions.

Term GPA refers to the grade point average of a student, on a 4.0 scale, for the semester they participated in the study. Cumulative GPA refers to the grade point average of a student, on a 4.0 scale, for the classes taken in all semesters prior to the semester they participated in the study. All the students in this data set were freshmen, hence, the term GPAs all correspond to the spring semester and the cumulative GPAs are based on the fall semester. GPA is a weighted average of the classes that a student took either in a given semester (term GPA) or in a series of semesters (cumulative GPA) with an A having the highest weight. We note that since the study only has freshman students as participants, our data is limited in its generalizability across a general student body which consists of non first-year students as well. The total sleep time is an average of how long a student slept each night with the exception of naps. In our original data set, total sleep time is reported in minutes, however, for better intuition, we have transformed the variable to be measured in hours.

Prior to our analysis, we will first explore the variables we are interested in. Let us begin by observing their distributions.
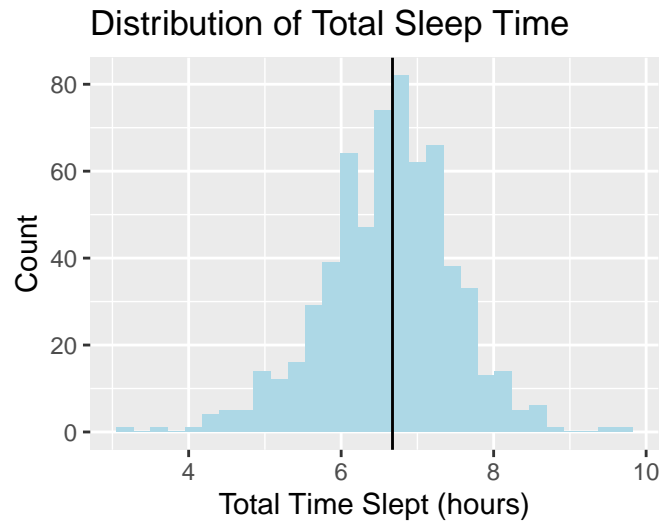
Distribution of Total Sleep Time

Figure 1

Figure one displays a histogram of the distribution of total sleep time with a vertical line representing the median at 6.6 hours. Total sleep time is a continuous variable and we can observe that it has a unimodal, bell-curve shaped, and somewhat symmetrical distribution about its median. The sleep times range from 0 hours to 10 hours with outliers lying between 0-4 hours and between 9-10 hours.
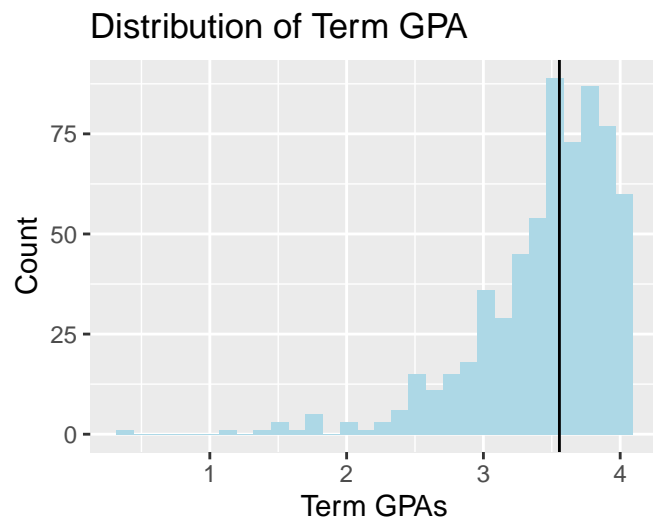
Distribution of Term GPA

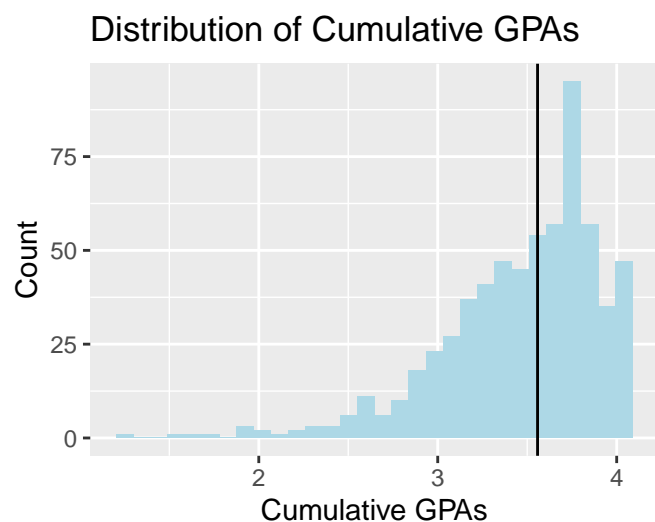Figure 2

Distribution of Cumulative GPAs

Figure 3

Figure two displays a histogram of the distribution of term GPAs and a vertical line representing the median around 3.51. Figures three also displays a histogram of the distribution of cumulative GPAs and a vertical line representing the median around 3.55. The distributions are similar since they are both unimodal and asymmetrical with a strong left skew. The term and cumulative GPA values range from 0-4, with outliers between 0-2. Most of the data is clustered around the median between a 3.0-4.0. We note that we performed transformations (e.g. log and square root) to obtain a bell-curve-like shaped distribution for the two variables, however, such transformations only scaled or shifted the distributions rather than transforming the shape. We also note that this skewness could have an an impact on the regression models that we will fit later in the methods section.

Next, we can look at scatter plots of the variables to see if there are any potential relationships we can build models for.
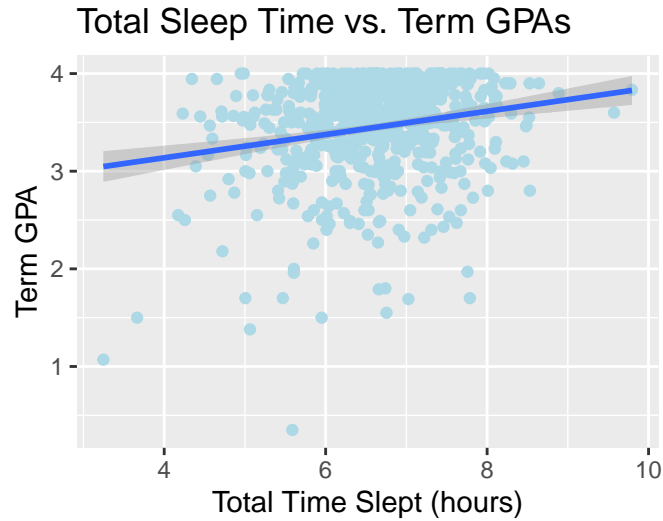
Total Sleep Time vs. Term GPAs

Figure 4

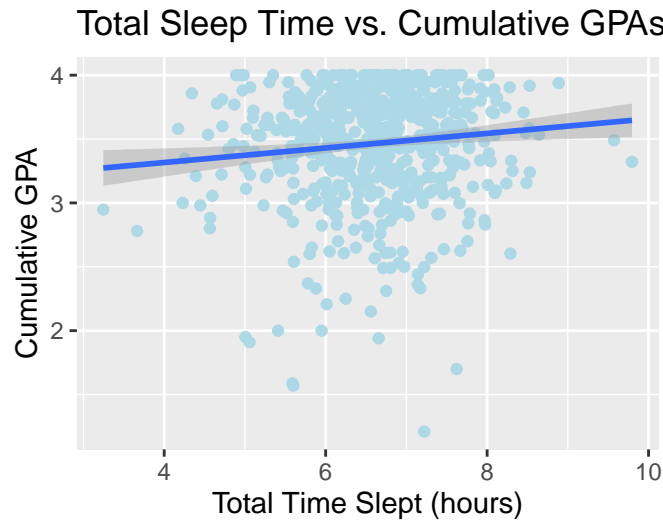## Total Sleep Time vs. Cumulative GPAs

Figure 5

Figure six and seven both display scatter plots of total time slept versus term GPA and cumulative GPA, respectively. We notice that both plots show data that are also clustered in the 5-8 hours range and GPA range of 3-4. There appears to be a positive but weak linear relationship displayed by the regression line for both plots.

Since total sleep time has a bell-curve distribution shape, we could do a log transformation of thr term and cumulative GPA to help make our models better. However, we note that doing so only clusters the data more rather than spreading it out. Thus, for our regression models, we will not have any variable transformations but we will note this as a limitation.

3. Methods

3.1 Research Question 1

To best investigate if sleep is associated with GPA, we can look at the relationship between sleep and term GPA as well as the relationship between sleep and cumulative GPA by fitting simple linear regression models for these relations respectively. We will use total sleep time as the predictor and term GPA and cumulative GPA will serve as the response variables for each respective model. Before we fit this model, we must remember that in our exploratory data analysis we transformed the sleep time to be in terms of hours and we also found skewness in the distributions of the GPA variables. This skewness could impact our models since we assume that the response variables are from a Normal Distribution for simple linear regression, which may not be the case for our daya. From here on out, for clarity's sake, model 1 will correspond to the linear model of Term GPA ~ Total Sleep Time and model 2 will correspond to the linear model of Cumulative GPA ~ Total Sleep Time.

We can determine if sleep is associated with GPA by performing t-tests (for 633 degrees of freedom) with the null hypothesis that $\hat{\beta}_1 = 0$. If there is an association between a particular type of GPA and sleep, we expect that for the respective model, $\hat{\beta}_1 \neq 0$. A key assumption for this testing is that the errors are normal, thus, $\hat{\beta}_1$ and $\hat{\beta}_0$ also follow the normal distribution. We also assume that the data are independent and randomly sampled, which is true in this case when looking at the population of university students that are first year students. Additionally, to define the nature of the association between sleep and GPA we can construct 95% confidence intervals for the $\hat{\beta}_1$'s from both models. We note that we will not test if $\hat{\beta}_0 = 0$ for each model since this does not correspond to a physically useful quantity–no student will survive a semester with an average of zero hours of sleep per night.

3.2 Research Question 2

To answer the second research question regarding the effect on average GPA if students get two hours less sleep on average, we can transform our total sleep time variable by subtracting two hours from each student's sleep time and then using this new data set to construct two new models–model 3 which corresponds to model 1 (Term GPA ~ Total Sleep Time) and model 4 which corresponds to model 2 (Cumulative GPA ~ Total Sleep Time). Comparing the parameter coefficients from the new models to the old models respectively will allow us

to observe if there are potentially any changes in the relationship between sleep and term GPA.

After fitting model 1 and model 2, we can construct a residual vs. fit plot for each model to determine if there are any violations of key assumptions.

Residuals vs. Fits: Model 1

Residuals

Figure 6

Residuals vs. Fits: Model 2
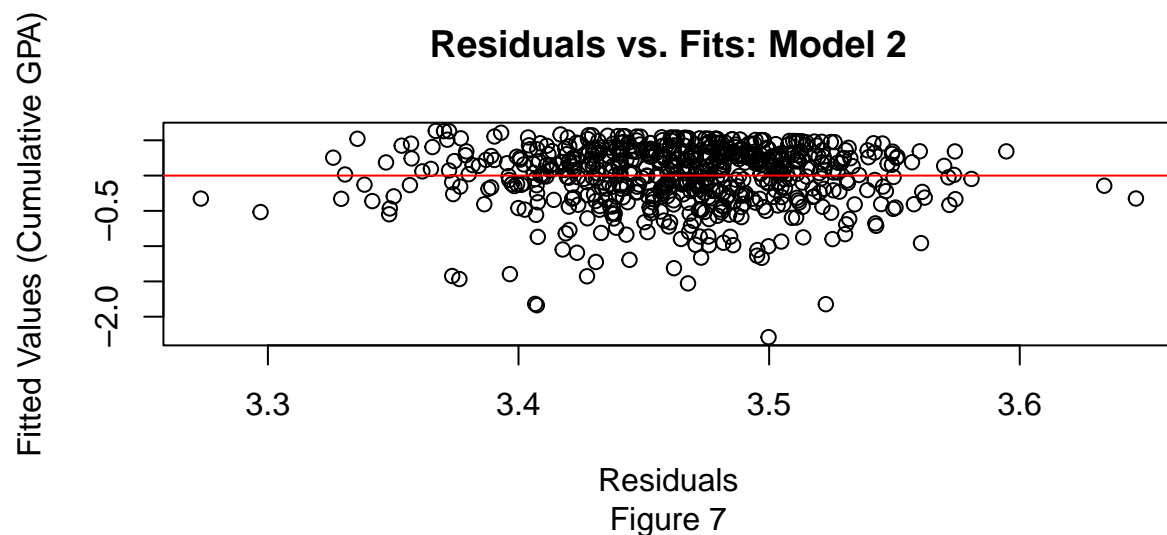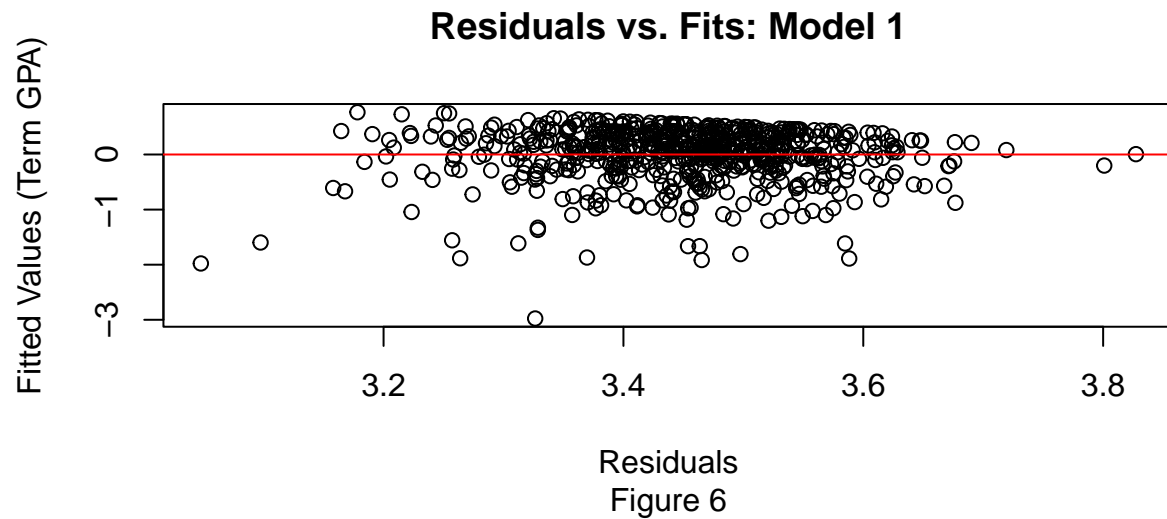
Residuals

Figure 7

In both plots from figure six and figure seven, we can observe that the residuals are not randomly scattered about the line y = 0. The residuals are also clustered towards the middle of the plot for fits between 3.4-3.5. This indicates that the residuals may not be uncorrelated, a violation of an assumption of the simple linear regression model. However, since we saw a potentially linear association in the scatter plots from figures four and five, we will proceed to fit our linear regression models with the note of this limitation.

We can also observe if our assumptions for the *t*-tests are potentially violated with a Q-Q

plot for the residuals of model 1 and model 2 against the normal distribution.

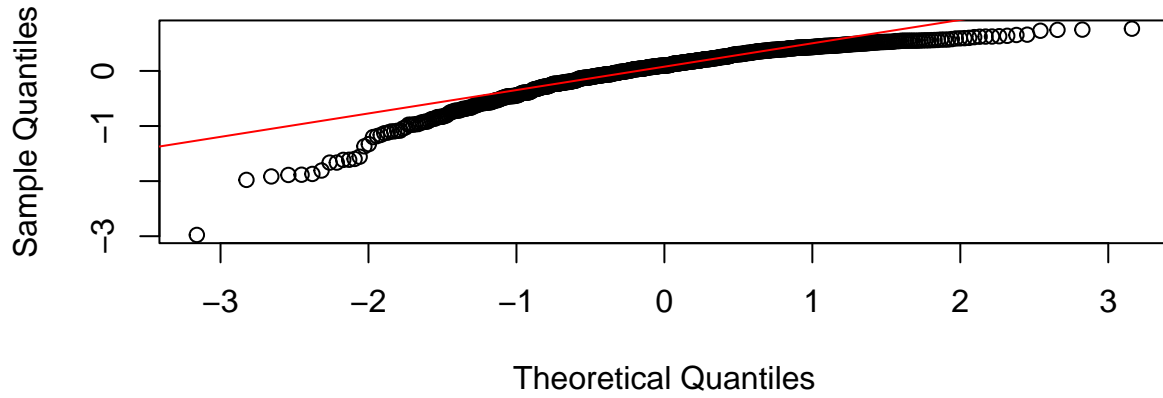**Figure 8 Normal Q–Q: Residuals From Model 1**

Figure 9 Normal Q–Q: Residuals From Model 2
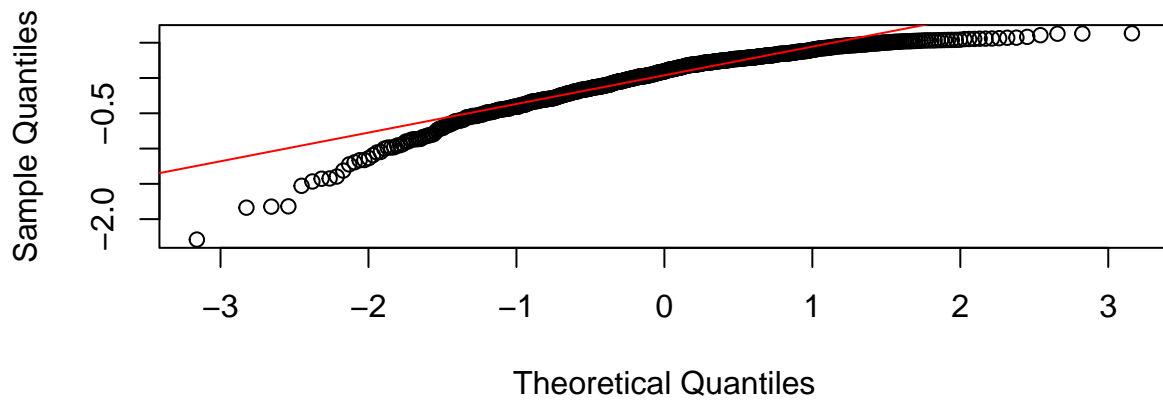
We can see that in both figures eight and nine, the residuals take on more extreme values than what we expect from a normal distribution, i.e. the distributions of both residuals are slightly heavy tailed so they may be non-normal. Since our t-test assumes the errors are normal to justify that the beta parameters also follow a normal distribution, we note this as a limitation for our confidence intervals construction.

3.3 Research Question 3

Finally, to investigate if there are other potential reasons for the relationship we observe between sleep and GPA, we can construct a scatter plot for term GPA versus cumulative GPA to observe the relationship between these two types of GPAs since we are using sleep as a predictor for both.

4. Results

4.1 Research Question 1

In this section, we will explore the outcomes, of the models we constructed in the previous section to answer our research questions. We will first analyze model 1, the simple linear regression function of Total Sleep Time ~ Term GPA.

Table 1: Coefficients for Model 1

term	estimate	std.error	statistic	p.value
(Intercept)	2.6611	0.1536	17.3274	0
TotalSleepTime	0.1191	0.0230	5.1765	0

Table 1 displays our model 1 coefficients as well as their corresponding standard errors, t statistics, and p-values. For this model, our regression function is $r(Total\widehat{Sleep}Time) = \hat{\beta}_1 * X + \hat{\beta}_0$ with term GPA as the response variable, $\hat{Y}$. A t-test for the null hypothesis that $\hat{\beta}_1 = 0$ results in $p = 0$. We reject the null hypothesis and conclude that sleep is positively associated with term GPA, $\hat{\beta}_1 = 0.1191$, $t(633) = 5.1765$, $p < 0.001$. One more hour of sleep at night is associated with an increase in term GPA by 0.1191 units, on average (95% CI [0.0739, 0.1642]). This model has an estimated variance of 0.2407 and a residual squared error of 0.4907.

Table 2: Coefficients for Model 2

term	estimate	std.error	statistic	p.value
(Intercept)	3.0883	0.1363	22.6653	0.0000
TotalSleepTime	0.0570	0.0204	2.7918	0.0054

Table 2 displays our model 2 coefficients as well as their corresponding standard errors, t statistics, and p-values. For this model, our regression function is again $r(Total\hat{Sleep}Time) = \hat{\beta}_1 * X + \hat{\beta}_0$ but with cumulative GPA as the response variable, $\hat{Y}$. A t-test for the null hypothesis that $\hat{\beta}_1 = 0$ results in $p = 0.0054$. We reject the null hypothesis and conclude that sleep is positively associated with cumulative GPA, $\hat{\beta}_1 = 0.0570$, $t(633) = 0.02041$, $p < 0.01$. One more hour of sleep at night is associated with an increase in cumulative GPA by 0.0054 units, on average (95% CI [0.0169, 0.0971]). For both model 1 and model 2 we see that students who sleep more may have higher term and cumulative GPAs and students who sleep for less time may have lower term and cumulative GPAs.

4.2 Research Question 2

Next, we will proceed to answer our second research question. We will analyze models 3 and 4 similar to how we analyzed models 1 and 2, then we will compare and contrast our parameter estimates between the models.

Table 3: Coefficients for Model 3

term	estimate	std.error	statistic	p.value
(Intercept)	2.8992	0.1081	26.8206	0
TotalSleepTime	0.1191	0.0230	5.1765	0

For model 3, a t-test for the null hypothesis that $\hat{\beta}_1 = 0$ results in $p = 0$. We reject the null hypothesis and conclude that sleep is positively associated with term GPA, $\hat{\beta}_1 = 0.1191$, $t(633) = 5.1765$, $p < 0.001$. One more hour of sleep at night is associated with an increase in term GPA by 0.1191 units, on average (95% CI [0.0739, 0.1642]). This is the same result as model 1, however, the intercept term is different between the two models. $\hat{\beta}_0 = 0.28992$ with a 95% confidence interval of [2.6869, 3.1112] for model 3, whereas $\hat{\beta}_0 = 2.6611$ with a 95% confidence interval [2.3595, 2.9636] for model 1. Since the two confidence intervals overlap, there is no significant difference in the $\hat{\beta}_0$ intercept coefficient between the two models. We find that the association between sleep and term GPA does not change for those who sleep 2 hours less.

Table 4: Coefficients for Model 4

term	estimate	std.error	statistic	p.value
(Intercept)	3.2022	0.0959	33.3893	0.0000
TotalSleepTime	0.0570	0.0204	2.7918	0.0054

For model 4, a t-test for the null hypothesis that $\hat{\beta}_1 = 0$ results in $p = 0.0054$. We reject the null hypothesis and conclude that sleep is positively associated with term GPA, $\hat{\beta}_1 = 0.0570$, $t(633) = 2.7918$, $p < 0.001$. One more hour of sleep at night is associated with an increase in cumulative GPA by 0.0570 units, on average (95% CI [0.0169, 0.0971]). This is the same result as model 2, however, again we see that the intercept term is different between the two models. $\hat{\beta}_0 = 3.2022$ with a 95% confidence interval of [3.0139, 3.3906] for model 4, whereas $\hat{\beta}_0 = 3.088$ with a 95% confidence interval [2.8207, 3.3558] for model 2. Since the two confidence intervals overlap, there is no significant difference in the $\hat{\beta}_0$ intercept coefficient between the two models. We find that the association between sleep and cumulative GPA does not change for those who sleep 2 hours less.
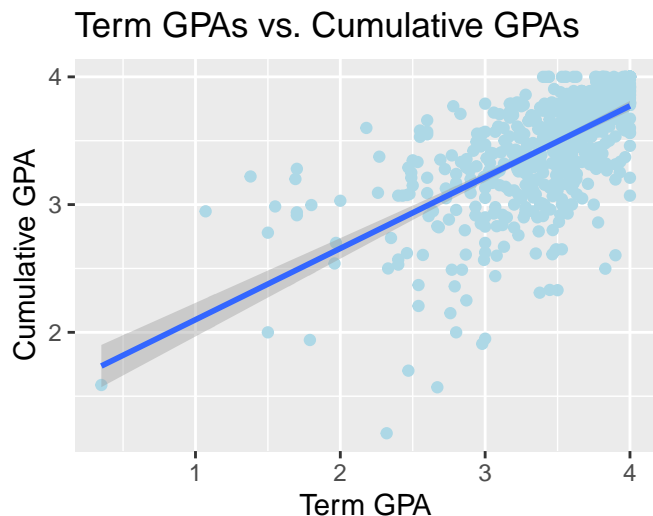
4.3 Research Question 3

Figure 10

Figure ten displays the relationship between term GPA and cumulative . We can observe that the data are clustered in the higher ranges of the GPA scale for each variable and that there seems to be a positive linear relationship between the two GPA types. Although we will

11

not explore this in further detail, for the sake of brevity, we will note that the relationship between term GPA and cumulative GPA may influence our findings for the relationships between sleep and term GPA and sleep and cumulative GPA since it may be the case that both types of GPAs are related.

5. Conclusion

In short, we found that sleep has a weak positive linear association with GPA. In particular, students who sleep less may have lower term and cumulative GPAs. With this said, the relationship between sleep and GPA may be influenced by other factors such as the relationship between term GPA and cumulative GPA themselves. The association between sleep and GPA remains the same even for students with two hours less of sleep on average, i.e. there is still a positive linear association. We recommend that the Vice Provost should not have classes starting at 8 am so students can get more sleep and perform better in their classes. However, we must note that there are several limitations in the generalizability of our analysis. First, we only had information on first-year students which means our study is not reflective of general trends across a university that has students across multiple year levels. Furthermore, we do not know which students in the study had 8 am classes, so we can not conclude if sleep has a direct impact on GPA for students who have 8 am classes. Additionally, the study was done at CMU and two other universities so the results from this report can not be applied to universities that did not conduct this study. Lastly, we can only apply our predictions for question two for those who do end up sleeping two hours less than typical, not those who sleep for less time by a different amount of hours.

We could improve our analysis in a variety of ways. To begin, the researchers can recruit more students for the study. They should track each student's average sleep time per night for the entirety of the study and also include students who are not first years. This will allow for a more comprehensive data set for a stronger analysis. We can also look at oddities in our data and adjust our models accordingly to account for these outliers which our initial analysis does not do.

Future research can be done to examine the effect of demographic factors like race, ethnicity, socio-economic status, etc…, as well as the number of courses/units being taken in a semester on term and cumulative GPA. These factors are elements that may influence a student's academic background coming into university and may continue to have an impact on GPA that is worth further exploring.