Question 1

Data was collected after the end of an introductory statistics course in order to investigate the relationship between certain variables, such as study time per week, and overall GPA. The table presented below shows the data for 8 females in the class on these variables.

(a) Create a scatter plot between study time and college GPA

(b) Analyze the data and report the equation and slope of the regression line

(c) Interpret the results – what information should be given to students concerning the impact of study time on academic performance?

(d) Find the predicted GPA for a student who spends 25 hours a week studying

student	study.time	missed.lectures	sleep.time	gpa
1	14	9	6.3	2.8
2	25	0	7.8	3.6
3	15	2	7.6	3.4
4	5	5	6.8	3.0
5	10	3	6.7	3.1
6	12	2	7.2	3.3
7	5	12	6.1	2.7
8	21	1	8.1	3.8

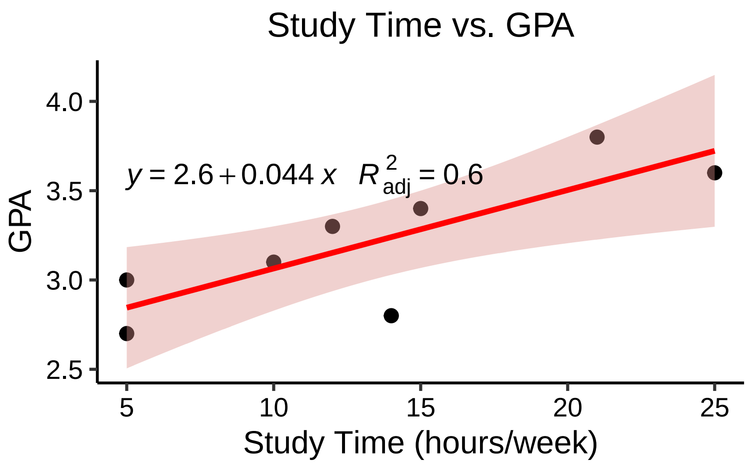
```r
# Fit the data to a linear model
fit <- lm(gpa ~ study.time, data=df)
summary(fit)
```

```

Call:
lm(formula = gpa ~ study.time, data = df)

Residuals:
     Min       1Q   Median       3Q      Max 
-0.43994 -0.12839  0.07592  0.14971  0.25270 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept)  2.62522    0.19381   13.54    1e-05 ***
study.time   0.04391    0.01299    3.38   0.0149 *  
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.243 on 6 degrees of freedom
Multiple R-squared:  0.6557,	Adjusted R-squared:  0.5983 
F-statistic: 11.43 on 1 and 6 DF,  p-value: 0.01485
```

Based on the results of the analysis, there appears to be an increase in the overall GPA of students as their weekly study time increases $\left(R_{adj} = 0.60, p = 0.015\right)$; specifically, for every additional hour of study time, there is an increase in GPA of 0.04.
The students should be informed that studying for a greater amount of time per week tends to increase their overall GPA.

```r
# Predict the GPA for a student who spends 25 hours a week studying
pred(fit, "a study time of 25 hours", study.time = 25)
```

```
[1] "For a student with a study time of 25 hours, GPA = 3.72"
```



Question 2

Repeat a-c from above with number of lectures missed as the predictive, x, variable. For part d, find the predicted GPA for a student who missed 12 lectures.

```r
# Fit the data to a linear model
fit <- lm(gpa ~ missed.lectures, data=df)
summary(fit)
```

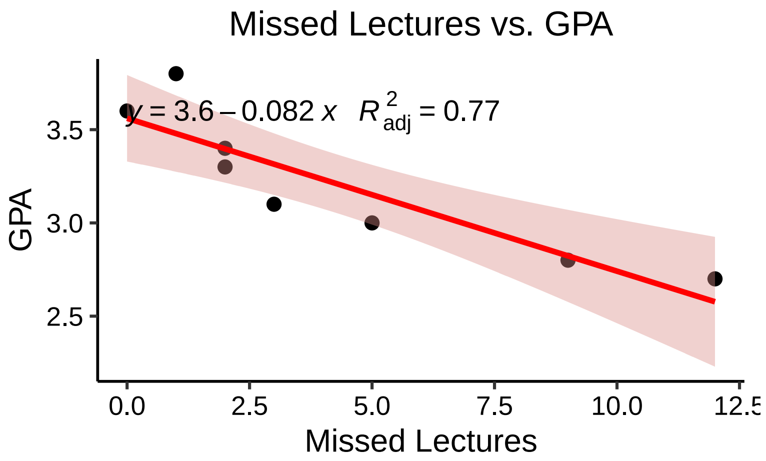
```
Call:
lm(formula = gpa ~ missed.lectures, data = df)

Residuals:
     Min       1Q   Median       3Q      Max 
-0.21498 -0.11048 -0.01002  0.06002  0.32105 

Coefficients:
                Estimate Std. Error t value Pr(>|t|)    
(Intercept)      3.56093    0.09477  37.574 2.37e-08 ***
missed.lectures -0.08198    0.01637  -5.007  0.00244 ** 
```

```
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.182 on 6 degrees of freedom
Multiple R-squared:  0.8069,	Adjusted R-squared:  0.7747 
F-statistic: 25.07 on 1 and 6 DF,  p-value: 0.002435
```

Based on the results of the analysis, there appears to be a decrease in overall GPA of students as the number of lectures they miss increases $\left(R_{adj} = 0.77, p = 0.002\right)$; specifically, for every lecture missed, there students' GPA decreases by 0.082.
The students should be informed that missing lectures tends to decrease their overall GPA.

```r
# Predict the GPA for a student who missed 12 lectures
pred(fit, "12 missed lectures", missed.lectures = 12)
```

```
[1] "For a student with 12 missed lectures, GPA = 2.58"
```



Question 3

Repeat a-c from above with average daily sleep time as the predictive, x, variable. For part d, find the predicted GPA for a student who averaged 7.8 hours of sleep.

```r
# Fit the data to a linear model
fit <- lm(gpa ~ sleep.time, data=df)
summary(fit)
```

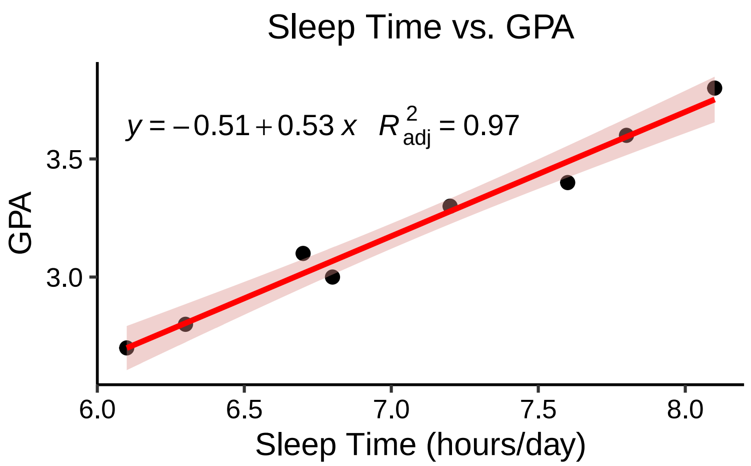
```
Call:
lm(formula = gpa ~ sleep.time, data = df)

Residuals:
      Min        1Q    Median        3Q       Max 
```

```
-0.088721 -0.020512  0.003267  0.028353  0.084801

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) -0.50990    0.22836  -2.233    0.067 .
sleep.time   0.52613    0.03213  16.374  3.3e-06 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.06126 on 6 degrees of freedom
Multiple R-squared:  0.9781,    Adjusted R-squared:  0.9745
F-statistic: 268.1 on 1 and 6 DF,  p-value: 3.304e-06
```

Based on the results, there appears to be an increase in overall GPA with an increase in the number of hours spent sleeping $\left(R_{adj} = 0.97, p = 3.3 \cdot 10^{-6}\right)$; specifically, the closer a student's sleep time is to 8 hours, the higher their GPA tends to be. Furthermore, average daily sleep time was found to be responsible for 97% of the variation in GPA.
The students should be informed that students who sleep for closer to 8 hours per day tend to have higher GPAs.

```
# Predict the GPA for a student who missed 12 lectures
pred(fit, "an average of 7.8 hours of sleep each day", sleep.time = 7.8)
```

```
[1] "For a student with an average of 7.8 hours of sleep each day, GPA = 3.59"
```



Question 4

Consider the study as described. Would this receive IRB approval? If not, why not and how could the study be changed so as to make the study more likely to be approved?

I think it is unlikely that the study would receive IRB approval as it is currently designed. For one, the study only includes 8 female students who were all enrolled in an introductory statistics course. Also, The data collected was from a convenience samlple, which is not representative of the population of students at the university. In order to improve the study, I would suggest that the researchers collect a stratified random sample of students from a variety of different majors, including multiple genders. In addition, the researchers would have to more clearly define how study time and average daily sleep time are calculated for each student.

Question 5

The above study was observational. Choose one of the factors and create an experimental study. Would your design be likely to receive IRB approval? What might some of the confounding factors be in your study design and how might you control for them?

Experimental Design

- **Sampling:** Take a stratified random sample of 100 students in their freshman year (all genders included) from the College of Engineering at Drexel University who are currently enrolled in classes. This should be done at the start of an academic quarter. Students should be stratified by gender.
  - The selected students should all be enrolled in an introductory statistics course which meets twice a week for 1 hour each.
  - The students should be randomly assigned to five groups of 20 students each.
    - Group 1: Students will be instructed to attend 100% of the lectures throughout the quarter.
    - Group 2: Students will be instructed to attend 75% of the lectures throughout the quarter.
    - Group 3: Students will be instructed to attend 50% of the lectures throughout the quarter.
    - Group 4: Students will be instructed to attend 25% of the lectures throughout the quarter.
    - Group 5: Students will be instructed not to attend any lectures throughout the quarter.
    - Students will be allowed to select which lectures they attend.
  - *Sample Size*, $n = 20 \ \forall$ groups
- **Measurements:** The students performance in the course will be measured based on their performance on the midterm and final exams. The midterm and final will be weighted equally in the final grade. The midterm will cover material from weeks 1-5 and the final will cover material from weeks 6-10. Students overall grade will be calculated as the average of their midterm and final exam scores.
- **Analysis:** The students' performance in the course will be compared across the five groups. The students' performance will be measured by their overall grade in the course. The data will be analyzed using a two-way ANOVA with the factors being number of lectures attendend and gender. If the ANOVA indicates that there is a significant difference in the students' performance across the five groups, then a post-hoc test will be performed to determine which groups are significantly different from each other.
- **Note:** This study will narrow the scope of the original study by only focusing on the effect of lecture attendance on the students' performance in a single course, instead of on their overall GPA.

IRB Approval
This study is unlikely to be approved by an IRB because there may not be an ethical way to fully control the sleep time of the students. Whereas study time could theoretically be controlled by the researchers, it would be much harder to require students to sleep for a certain amount of time each day throughout an entire

quarter; depending on the nature of their courseload and extracurricular activities, students are likely to have varying amounts of sleep. Furthermore, it would likely be an invasion of the students' privacy to implement any strict monitoring system for their sleep schedules throughout the 10 weeks of the term. Additionally, since the study is only being conducted on students in the College of Engineering taking an introductory statistics course, the results of the study may not be generalizable to the entire student population at Drexel University. Finally, since the study could have a negative impact on the students' GPA which could have consequences on their emotional and mental health, and may impact their job prospects since GPA is often used as a screening tool by employers.
Confounding Factors

- Students may have varying levels of prior experience with the material covered in the course
- The amount of time spent studying for the course may vary across the groups which would likely impact their performance in the course. To control for this, students across all groups should be required to study for an average of 12 hours per week. For the purposes of this study, "study time" will be defined as the number of hours spent working on course material outside of class time.
- The amount of sleep time (as shown in our analyses above) has a significant impact on students' overall GPA. To control for this, students across all groups should be required to sleep for an average of 8 hours per day.
- There students' performance could also be impacted by variations in course instructors, teaching assistants, and the lecture material used. To control for this, all students in the study should be taught the same material by the same instructor and teaching assistants throughout the quarter.

Question 6

Run the data table as a multiple regression using all three predictive variables and GPA as the variable being predicted. Do the conclusions change? How?

```r
# Fit the data to a linear model
fit <- lm(gpa ~ sleep.time + study.time + missed.lectures, data=df)
summary(fit)
```

```
Call:
lm(formula = gpa ~ sleep.time + study.time + missed.lectures,
    data = df)

Residuals:
       1        2        3        4        5        6        7        8
-0.02854 -0.02630 -0.07300 -0.04530  0.05702  0.01755  0.02951  0.06905

Coefficients:
                 Estimate Std. Error t value Pr(>|t|)
(Intercept)      0.065423   0.675159   0.097  0.92747
sleep.time       0.441711   0.094572   4.671  0.00951 **
study.time       0.004621   0.005911   0.782  0.47806
```

```
missed.lectures -0.009373   0.013358  -0.702  0.52155
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.06724 on 4 degrees of freedom
Multiple R-squared:  0.9824,    Adjusted R-squared:  0.9692
F-statistic: 74.51 on 3 and 4 DF,  p-value: 0.000576
```

When we look at a multiple regression model, only daily sleep time was found to be a significant predictor of GPA. This is because the other variables were likely confounding variables and the correlation between them and GPA was spurious.