Sleep Cycle Analysis

Jonathan Chia, Krystine Osumo, Gregory Pollock

Introduction

The purpose of our experiment was to gain information about what affects sleep quality using the Sleep Cycle application's measure of sleep quality. Because we conducted an observational study, we can only conclude correlation and assume that Jonathan's data is representative of a typical college student. We looked at the effects of total time slept and the season of the year on sleep quality. The principle questions of interest were: "Does the amount you sleep affect your overall sleep quality?" and "Does the season of the year affect your overall sleep quality?" Our initial assumptions were that Sleep quality will increase when you sleep more, student will have better sleep quality in spring/summer compared to fall/winter, there may not be a difference between 8-9 hours of sleep versus more than 9 hours, and that there will be an interaction between season and amount of sleep.

Design and Data Collection

Null Hypothesis / Alternative Hypothesis

Our three null hypotheses assume: that there is no difference in mean sleep quality between Fall, Winter and Spring / Summer seasons, that there is no difference in mean sleep quality when sleeping between 5 - 7.25 hours, 7.25 - 8 hours , 8 - 9 hours and sleeping more than 9 hours, and that there is no interaction between the seasons and the amount of sleep.

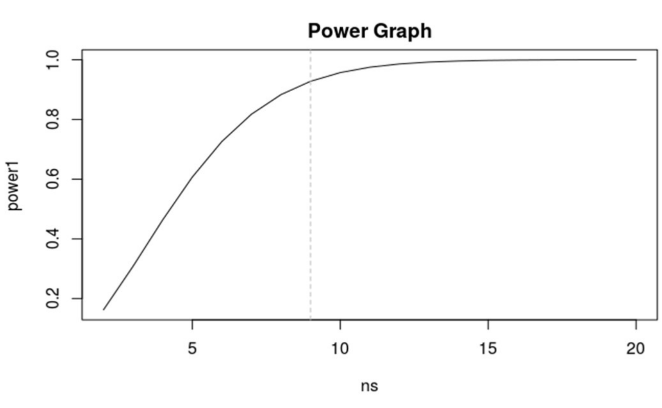
As our alternative hypothesis, we assume that at least one mean sleep quality measure differs from the rest of the seasons; we also assume that at least one mean sleep quality measure differs from the rest of the amount of sleep. We also assume that there is an interaction between the two factors.

Model

We performed a 2 factor Anova Test with Sleep Quality as our response variable. The two factors are: Hours Jonathan Slept (4 levels: 5 - 7.25 hrs, 7.25 - 8 hrs, 8-9 hrs, 9+ hrs) and Seasons (3 levels: Fall, Spring and Summer, Winter).

Below is our model where Y_{ijk} represents the observation, μ represents the grand average of all levels, α_i represents the season effect with i level, β_j represents the amount of sleep effect with j levels, $\alpha_i\beta_j$ represents the interaction effect between season effect and amount of sleep effect and ε_{ijk} represents the effect of random error.

$$Y_{ijk} = \mu + \alpha_i + \beta_j + \alpha_i\beta_j + \varepsilon_{ijk}$$



Because this is an observational study, we used power analyses to find what differences in sleep quality are statistically possible to detect. The within.variance is the variance of the response variable, the sleep quality. The between variance was calculated assuming that we want to detect a difference of at least 2.5% between Spring/Summer and Fall/Winter as well as detect a difference of at least 5% between each level of time sleeping. With at least 9 replications in each cell, we can detect a difference of 2.5% in sleep quality with a power of about 90%.

Data Collection

Data was gathered through the Sleep Cycle iPhone application. The user is instructed to place the phone on a bedside table while the app measures movement with a microphone to provide a measure of sleep quality with high movement or irregular movement patterns suggesting bad sleep. We used the *Time in Bed*, *Sleep Quality* and *Date* measurements from the app to gather our data. We also created season and time sleeping (timslpfac) factors. A sample of our data is included in the appendix. Note that the data has limitations because one can decide when to sleep and wake up, so we have to assume the *Time in Bed* factor is random in order to do analysis.

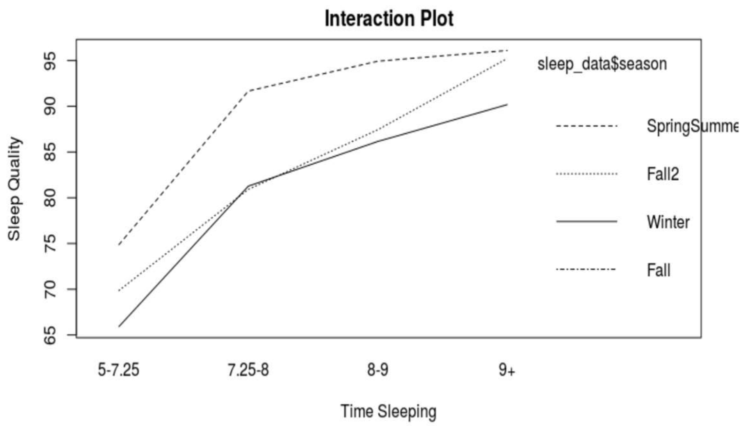
Data Analysis

After creating boxplots, we saw that a logarithmic transformation would be unnecessary since the largest variance divided by the smallest variance, was less than 2. Therefore, we concluded that they have equal variance. We also checked the normality of the observations and also the normality of residuals by looking at their Histograms. The results showed that the histogram of the observations was left skewed but did not show any extreme outliers. The histogram of the residuals was slightly left skewed but overall it resembled a normal curve. With the data satisfying the assumptions, we then performed the Analysis of Variance.

Since we had an unbalanced experiment, we used the Type III Anova Test. Figure 1.1 looks at Time Sleeping taking into account the Season effect and we see that the p-value for Time Sleeping is sufficiently small. Figure 1.2 does the same for Season taking into account the Time Sleeping effect. The new calculations, in Figure 1.2, show that with the time sleeping effect taken into account, we saw that the p-value for Season became smaller, while the p-value for Time Sleeping remained unchanged. With Time Sleeping being the most significant, Figure 1.2 is the most representative Anova table, and we can conclude that both factors are significant.

Figure 1.1	Degrees of Freedom	Sum of Squares	Mean of Squares	F Value	P-Value
Season	2	3364.7	1682.4	29.0993	2.775×10^{-12}
Time Sleeping	3	21842.8	7280.9	125.9363	$< 2.2 \times 10^{-16}$
Season	6	546.0	91.0	1.5741	0.1543
Residuals	302	17460.0	57.8		

Figure 1.2	Degrees of Freedom	Sum of Squares	Mean of Squares	F Value	P-Value
Time Sleeping	4	2241812	560453	9693.9970	$< 2.2 \times 10^{-16}$
Season	2	4527	2263	39.1487	7.628×10^{-16}
Interaction	6	5461	91	1.5741	0.1543
Residuals	302	17460	58		



While checking the Anova Table, we found that the p-value for interaction is 0.1543. Thus, we can conclude that while there is an interaction, it is not significant.

Figure 1.3 Confidence Intervals of Means

Amount of Sleep (hrs)	2.5%	97.5%
5-7.25	66.5920121	73.1222736
7.25 - 8	78.4228818	83.4104515
8-9	85.0399465	89.8318484
9+	90.2346526	100.2097919

The confidence interval support what the initial box plot showed. There appears to be a linear relationship between time sleeping and sleep quality. The means of the levels of Amount of Sleep increase as the amount of sleep increase.

Conclusions

The data suggest sleep quality improves in a somewhat linear fashion. We concluded that Season has an effect on overall sleep quality and reject that null hypothesis based on a p-value of 2.775×10^{-12}. We also concluded that Time Sleeping has an effect on overall sleep quality based on a p-value of 2.2×10^{-16}. There is an interaction between Season and the Time Sleeping, but it is not significant (p-value = 0.1543). We can conclude that there is correlation for Time Sleeping and sleep quality and correlation for Season and sleep quality, but we are unable to generalize our findings to all college students due to our one-person population.

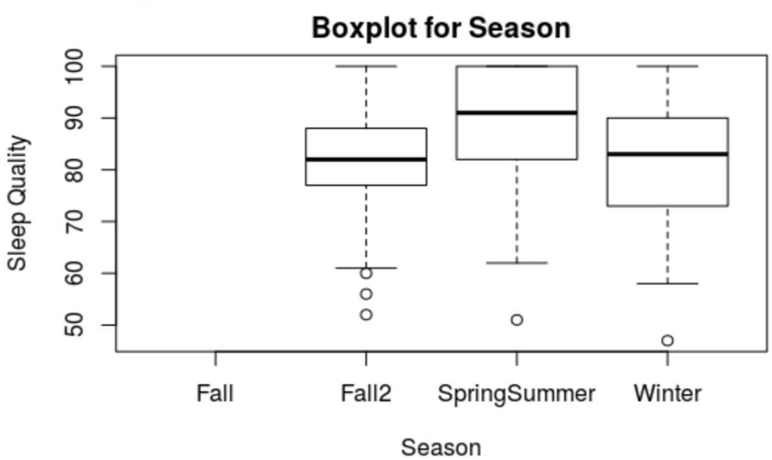
For this project, we have assumed that Jonathan represents a typical college student; however, everyone sleeps differently so these results are only catered to him. We were unable to measure some factors that are commonly known to affect sleep. These factors include stress, outside noises, diet or phone light before sleeping. Additionally, consistent sleep patterns probably have an effect on sleep quality, meaning the previous days' sleep can affect the next day. However, since we have a large sample data, we believe it is enough data to balance out the effects. If we had more time and resources, we could have more students use the Sleep Cycle App and report their data since a broader sample could better represent the population of college students. The best method would be to run a sleep study experiment, with randomized treatments from a random sample. Then we could conclude causation.

Appendix

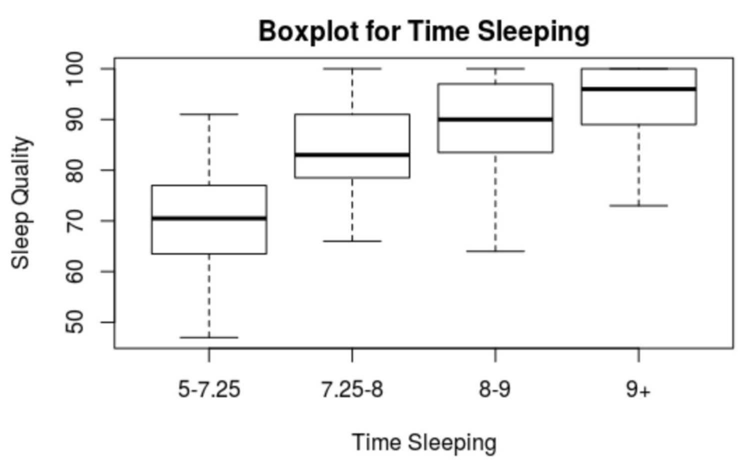
Data Sample

	Start	End	Sleep.quality	Time.in.bed	timesleeping	season	tmslpfac	Startf
163	2018-01-08 22:52:08	2018-01-09 12:30:56	73	13:38	818.8000	Winter	9+	Normal
164	2018-01-10 01:19:13	2018-01-10 06:50:15	47	5:31	331.0333	Winter	5-7.25	Late
166	2018-01-11 00:01:31	2018-01-11 08:49:37	92	8:48	528.1000	Winter	8-9	Late
167	2018-01-11 23:53:42	2018-01-12 08:50:29	90	8:56	536.7833	Winter	8-9	Normal
169	2018-01-13 00:22:34	2018-01-13 06:49:35	74	6:27	387.0167	Winter	5-7.25	Late
170	2018-01-14 00:00:11	2018-01-14 10:05:15	97	10:05	605.0667	Winter	9+	Late
171	2018-01-14 23:48:46	2018-01-15 09:05:19	79	9:16	556.5500	Winter	9+	Normal
172	2018-01-16 00:21:47	2018-01-16 07:59:32	77	7:37	457.7500	Winter	7.25-8	Late
173	2018-01-17 00:02:20	2018-01-17 06:45:38	82	6:43	403.3000	Winter	5-7.25	Late
175	2018-01-17 23:59:29	2018-01-18 08:29:20	83	8:29	509.8500	Winter	8-9	Late
178	2018-01-19 00:51:13	2018-01-19 08:30:32	89	7:39	459.3167	Winter	7.25-8	Late
179	2018-01-20 00:49:41	2018-01-20 06:49:18	71	5:59	359.6167	Winter	5-7.25	Late
180	2018-01-21 01:01:22	2018-01-21 09:30:25	94	8:29	509.0500	Winter	8-9	Late
181	2018-01-21 23:05:11	2018-01-22 06:50:50	84	7:45	465.6500	Winter	7.25-8	Normal
183	2018-01-23 00:09:56	2018-01-23 08:59:50	95	8:49	529.9000	Winter	8-9	Late

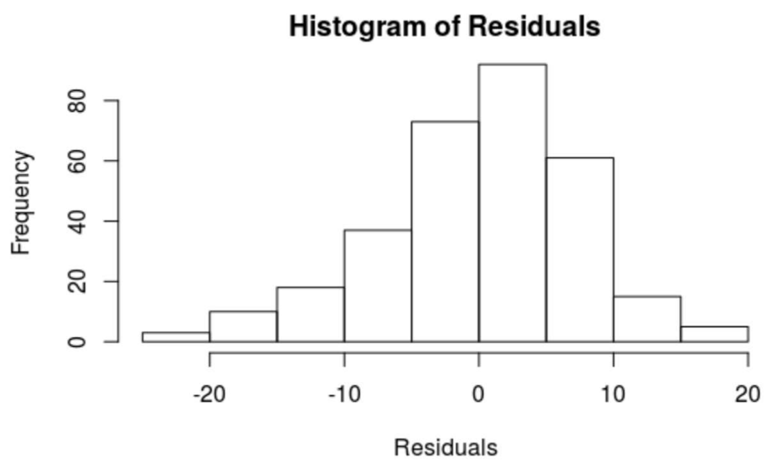
Extra Figures



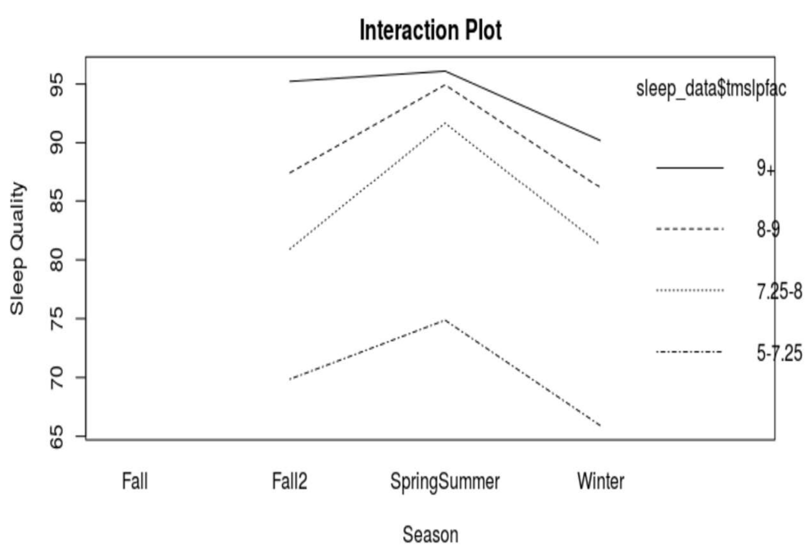
We started the data with two fall semesters, but we noticed that Jonathan's sleep quality was irregular for the first fall semester due to higher stress and slight insomnia. Therefore, we decided that for this study, it would be better to not take into account the first fall semester.



The Histogram for Sleep Quality (the response) does not show any significant outliers.



The Histogram of Residuals is approximately normal.



R Code

Read in data, clean it up

```
library(tidyverse)
full_data <- read.csv("/cloud/project/sleepdata.csv", sep = ";",stringsAsFactors = FALSE)
```

```
#remove empty variables by making subset dataframe, remove winter2 dates
sleep_data <- full_data[-c(1,574:587) ,1:4]
#make sure dates are coerced into "POSIX" variables not factors
sleep_data$Start <- as.POSIXlt(sleep_data$Start)
class(sleep_data$Start)
#do the the same for end
sleep_data$End <- as.POSIXlt(sleep_data$End)
length(sleep_data$Sleep.quality)
#format sleepquality into numeric
sleep_data$Sleep.quality <- gsub("%", "", sleep_data$Sleep.quality)
sleep_data$Sleep.quality <- as.numeric(sleep_data$Sleep.quality)
#format into minutes
sleep_data$timesleeping <- sleep_data$End-sleep_data$Start
#Removing 'naps' from data frame
sleep_data = sleep_data[!sleep_data$timesleeping < 300.00,]
```

Make Season Factor

```
getSeason <- function(DATES) {
Winter <- as.Date("2018-1-08", format = "%Y-%m-%d") # Winter Semester
SpringSummer <- as.Date("2018-5-01",  format = "%Y-%m-%d") # Spring+Summer Semester
Fall2 <- as.Date("2018-9-04",  format = "%Y-%m-%d") # Fall Semester
Fall <- as.Date("2017-9-05",  format = "%Y-%m-%d")
    ifelse(DATES < Winter, "Fall",
     ifelse (DATES >= Winter & DATES < SpringSummer, "Winter",
     ifelse (DATES >= SpringSummer & DATES < Fall2, "SpringSummer", "Fall2"))}
sleep_data$season <- as.factor(unlist(lapply(sleep_data$Start, getSeason)))
sleep_data <- sleep_data[which(sleep_data$season != "Fall"),]
```

Time sleeping to Factor

```
brk1 <- 300
brk2 <- 435
brk3 <- 480
brk4 <- 540
brk5 <- 1000
sleep_data$tmslpfac <- as.factor(cut(as.numeric(sleep_data$timesleeping), breaks = c(brk1, brk2,
brk3, brk4, brk5), labels = c("5-7.25", "7.25-8", "8-9", "9+"), right = TRUE)) #Makes time
sleeping a factor and introduces breaks to sort into our categories
```

Decomposition table

```
winter_5 <- length(sleep_data$Sleep.quality[sleep_data$season == "Winter" &
sleep_data$tmslpfac == "5-7.25"])
winter_7 <- length(sleep_data$Sleep.quality[sleep_data$season == "Winter" &
sleep_data$tmslpfac == "7.25-8"])
winter_8 <- length(sleep_data$Sleep.quality[sleep_data$season == "Winter" &
sleep_data$tmslpfac == "8-9"])
winter_9 <- length(sleep_data$Sleep.quality[sleep_data$season == "Winter" &
sleep_data$tmslpfac == "9+"])
SS_5 <- length(sleep_data$Sleep.quality[sleep_data$season == "SpringSummer" &
sleep_data$tmslpfac == "5-7.25"])
```

```
SS_7 <- length(sleep_data$Sleep.quality[sleep_data$season == "SpringSummer" &
sleep_data$tmslpfac == "7.25-8"])
SS_8 <- length(sleep_data$Sleep.quality[sleep_data$season == "SpringSummer" &
sleep_data$tmslpfac == "8-9"])
SS_9 <- length(sleep_data$Sleep.quality[sleep_data$season == "SpringSummer" &
sleep_data$tmslpfac == "9+"])
fall2_5 <- length(sleep_data$Sleep.quality[sleep_data$season == "Fall2" & sleep_data$tmslpfac
== "5-7.25"])
fall2_7 <- length(sleep_data$Sleep.quality[sleep_data$season == "Fall2" & sleep_data$tmslpfac
== "7.25-8"])
fall2_8 <- length(sleep_data$Sleep.quality[sleep_data$season == "Fall2" & sleep_data$tmslpfac
== "8-9"])
fall2_9 <- length(sleep_data$Sleep.quality[sleep_data$season == "Fall2" & sleep_data$tmslpfac
== "9+"])
```

Check for Outliers

```
hist(sleep_data$Sleep.quality, main = "Histogram of Sleep Quality", xlab = "Sleep Quality")
```

Boxplots to make sure variances are about equal

```
boxplot(sleep_data$Sleep.quality ~ sleep_data$season, data = sleep_data, main = "Boxplot for
Season", xlab = "Season", ylab = "Sleep Quality")
#log
#boxplot(log(sleep_data$Sleep.quality) ~ sleep_data$season, data = sleep_data, main = "Boxplot
for Season2", xlab = "Season", ylab = "Sleep Quality")
#For time intevrals
boxplot(sleep_data$Sleep.quality ~ sleep_data$tmslpfac, data = sleep_data,  main = "Boxplot for
Time Sleeping", xlab = "Time Sleeping", ylab = "Sleep Quality")
```

Interaction plots

```
#1. for season
Season <- interaction.plot(sleep_data$season, sleep_data$tmslpfac, response =
sleep_data$Sleep.quality, main = "Interaction Plot", xlab = "Season", ylab = "Sleep Quality")
#2. for time intevrals
Tmslpfac <- interaction.plot(sleep_data$tmslpfac, sleep_data$season, response =
sleep_data$Sleep.quality, main = "Interaction Plot", xlab = "Time Sleeping", ylab = "Sleep
Quality", legend = TRUE)
```

Compare means

```
library(dplyr)
x <- sleep_data %>% select(Sleep.quality, season, tmslpfac) %>% filter(season == "Winter",
tmslpfac == "5-7.25") %>% select(Sleep.quality)
 mean(x$Sleep.quality)
 var(x$Sleep.quality)
x2 <- sleep_data %>% select(Sleep.quality, season, tmslpfac) %>% filter(season == "Winter",
tmslpfac == "7.25-8") %>% select(Sleep.quality)
 mean(x2$Sleep.quality)
 var(x2$Sleep.quality)
x3 <- sleep_data %>% select(Sleep.quality, season, tmslpfac) %>% filter(season == "Winter",
tmslpfac == "8-9") %>% select(Sleep.quality)
 mean(x3$Sleep.quality)
 var(x3$Sleep.quality)
x4 <- sleep_data %>% select(Sleep.quality, season, tmslpfac) %>% filter(season ==
"SpringSummer", tmslpfac == "5-7.25") %>% select(Sleep.quality)
 mean(x4$Sleep.quality)
  var(x4$Sleep.quality)
```

```
x5 <- sleep_data %>% select(Sleep.quality, season, tmslpfac) %>% filter(season ==
"SpringSummer", tmslpfac == "7.25-8") %>% select(Sleep.quality)
 mean(x5$Sleep.quality)
 var(x5$Sleep.quality)
x6 <- sleep_data %>% select(Sleep.quality, season, tmslpfac) %>% filter(season ==
"SpringSummer", tmslpfac == "8-9") %>% select(Sleep.quality)
 mean(x6$Sleep.quality)
var(x6$Sleep.quality)
x7 <- sleep_data %>% select(Sleep.quality, season, tmslpfac) %>% filter(season ==
"SpringSummer", tmslpfac == "5-7.25") %>% select(Sleep.quality)
 mean(x7$Sleep.quality)
 var(x7$Sleep.quality)
 x8 <- sleep_data %>% select(Sleep.quality, season, tmslpfac) %>% filter(season ==
"SpringSummer", tmslpfac == "7.25-8") %>% select(Sleep.quality)
 mean(x8$Sleep.quality)
 var(x8$Sleep.quality)
x9 <- sleep_data %>% select(Sleep.quality, season, tmslpfac) %>% filter(season ==
"SpringSummer", tmslpfac == "8-9") %>% select(Sleep.quality)
 mean(x9$Sleep.quality)
  var(x9$Sleep.quality)
```

Power curve

```
ns <- 2:20
delta <- 2.5
power1 <- power.anova.test (groups = 12, ns, within.var = var(sleep_data$Sleep.quality),
between.var = var(c(0, delta, 0, 2*delta, 3*delta, 2*delta, 4*delta, 5*delta, 4*delta,6*delta,
7*delta, 6*delta)), sig.level = 0.05)$power
plot(ns,power1, type ="l", main = "Power Graph")
abline(v=9, lty=2, col = "gray")
```

Anova and CI

```
season <- as.factor(sleep_data$season)
tmslpfac <- as.factor(sleep_data$tmslpfac)
#make the linear model
model <- lm(Sleep.quality ~ season  + season:tmslpfac + tmslpfac, data = sleep_data)
model2<- lm(Sleep.quality ~ season:tmslpfac + tmslpfac + season, data = sleep_data)
#run anova ( type III)
anova(model)
anova(model2)
#make sure residuals are normal
hist(model2$residuals,main = "Histogram of Residuals", xlab = "Residuals")
#confidence intervals
model2 <- lm(Sleep.quality ~ -1 + tmslpfac + season + season:tmslpfac, data = sleep_data)
confint(model2)
```