2021 in Data

Krista ONeill

1/1/2022

## 2021 in Data

For a year, I logged my hourly activities as well as my mood for each day on a scale of -1 to -6.

The legend for the grading scale & the list of activities can be found in the attached file.

df <- read.csv('2021 in data.csv', header=TRUE)  
  
#viewing the column types and converting them all to character  
#sapply(df, class)  
  
df <- data.frame(lapply(df, as.character), stringsAsFactors=FALSE)  
df$Grade <- as.numeric(df$Grade)  
df$WorkHours <- as.numeric(df$WorkHours)  
df$Sleep <- as.numeric(df$Sleep)

#subsetting; finding basic statistic values  
sleep <- subset(df, Sleep >= 7)  
nosleep <- subset(df, Sleep < 7)  
  
#adding new column based on sleep values  
df$SleptEnough <- ifelse(df$Sleep >= 7, 1, 0)  
  
mean\_grade <- mean(df$Grade)  
cat("Average grade of 2021:", mean\_grade)

## Average grade of 2021: -4.243836

mean\_sleep <- mean(df$Sleep)  
cat("Average hours of sleep:", mean\_sleep)

## Average hours of sleep: 7.293151

mean\_work <- mean(df$WorkHours)  
cat("Average hours worked per day:", mean\_work)

## Average hours worked per day: 5.087671

We will now perform an independent two-sample T-test to see if there is a significant statistical difference in grade between the days I slept enough (>= 7 hrs) and the days I didn’t.

test1 = t.test(Grade ~ SleptEnough, var.equal=FALSE, data = df)  
  
test1

##   
## Welch Two Sample t-test  
##   
## data: Grade by SleptEnough  
## t = -2.8547, df = 349.17, p-value = 0.004565  
## alternative hypothesis: true difference in means between group 0 and group 1 is not equal to 0  
## 95 percent confidence interval:  
## -0.44777211 -0.08246045  
## sample estimates:  
## mean in group 0 mean in group 1   
## -4.400000 -4.134884

This t-test shows that the p-value is 0.004, meaning the difference is significant. The mean for Group 0 (“Didn’t sleep enough”) is -4.4, while the mean for Group 1 is ~ -4.135. While this doesn’t seem like a big difference, the scale goes only from -1 to -7.

#Exporting new csv for visualization.  
write.csv(df, "2021 in data sleptenough.csv", row.names=FALSE)