Final Project

Maida Ismail

2024-12-01

```
library(readxl)
library(ggplot2)
library(plotly)

library(tm)

library(tidyverse)
```

μ_{Yes}: Mean of sleep hours on the days I went to the gym.

μ_{No}: Mean of sleep hours on the days I did not go to the gym.

$H_0: \mu_{Yes} <= \mu_{No}$

$H_A: \mu_{Yes} > \mu_{No}$

```
url = "~/OneDrive - Minnesota State/Desktop/Data 211/FinalProjectData.xltx"
data = read_excel(url)

t.test(data$`Sleep hrs`~ data$`Go to the gym?`, alternative = "greater")
```

```
##
##  Welch Two Sample t-test
##
## data:  data$`Sleep hrs` by data$`Go to the gym?`
## t = -0.15586, df = 10.482, p-value = 0.5604
## alternative hypothesis: true difference in means between group No and grou
p Yes is greater than 0
## 95 percent confidence interval:
##  -1.796423       Inf
## sample estimates:
##  mean in group No mean in group Yes
##          6.214286          6.357143
```

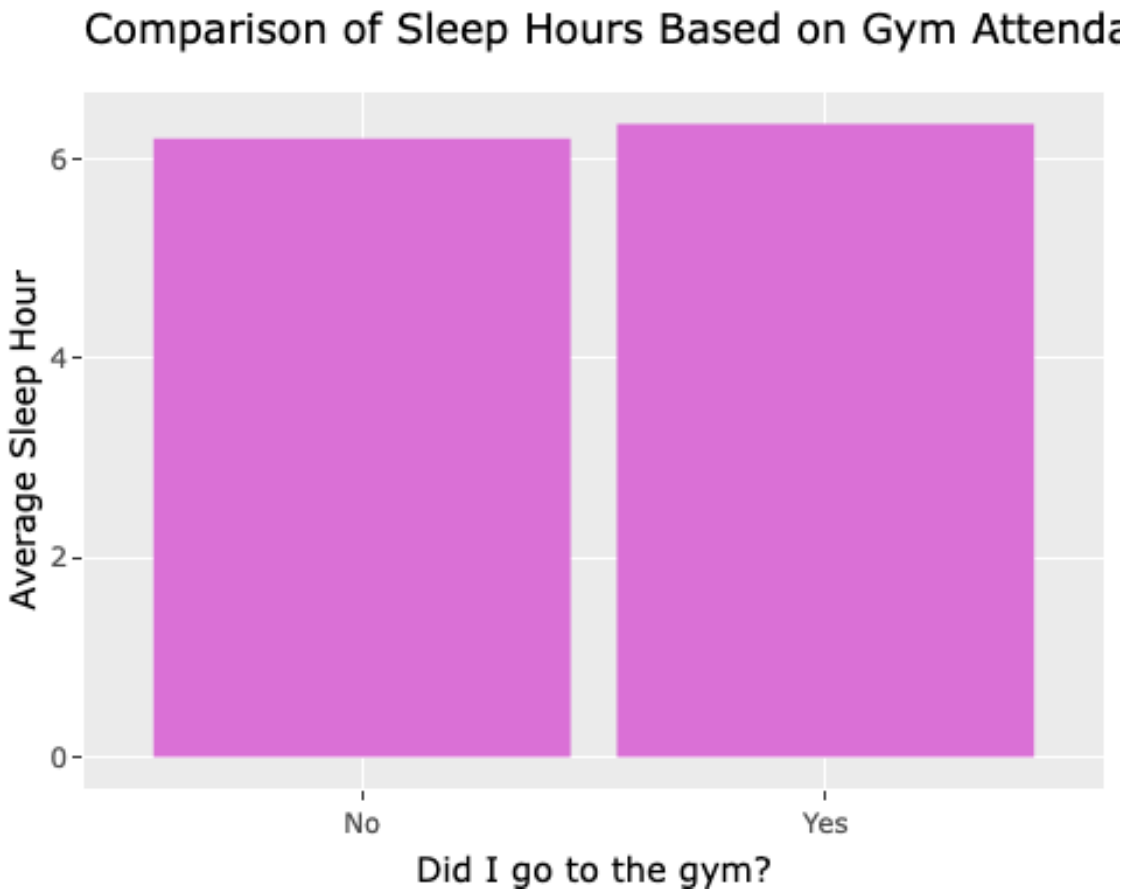
Since the p-value is 0.5604 which is greater than the significance level $\alpha = 0.05$, I fail to reject the null hypothesis. That means there is not enough evidence to support the claim that the average hours of sleep on the days I went to the gym is significantly higher than the days I did not visit the gym.

A bar graph to show the mean

```
mean_data = data %>%
  group_by(`Go to the gym?`) %>%
  summarize(
    mean_by_sleep = mean(`Sleep hrs`, na.rm = TRUE)
  )

graph = ggplot(mean_data, aes(x = `Go to the gym?`, y = mean_by_sleep)) +
  geom_bar(stat = "identity", fill = "orchid") +
  labs(x = "Did I go to the gym?", y = "Average Sleep Hour", title = "Compari
son of Sleep Hours Based on Gym Attendance")


ggplotly(graph)
```



A bar graph with error bars

```
mean_data = data %>%
  group_by(`Go to the gym?`) %>%
  summarize(
    mean_by_sleep = mean(`Sleep hrs`, na.rm = TRUE),
    se_sleep = sd(`Sleep hrs`, na.rm = TRUE) / sqrt(n())
  )

graph = ggplot(mean_data, aes(x = `Go to the gym?`, y = mean_by_sleep)) +
  geom_bar(stat = "identity", fill = "orchid") +
  labs(x = "Did I go to the gym?", y = "Average Sleep Hour", title = "Compari
son of Sleep Hours Based on Gym Attendance") +
  geom_errorbar(aes(ymin = mean_by_sleep - se_sleep, ymax = mean_by_sleep + s
e_sleep), width = 0.2)


ggplotly(graph)
```