Final Project

Marisol Arita

2023-11-27

```
library(tidyverse)
```

```
## ── Attaching core tidyverse packages ──────────────────────── tidyverse
2.0.0 ──
## ✓ dplyr     1.1.3     ✓ readr     2.1.4
## ✓ forcats   1.0.0     ✓ stringr   1.5.0
## ✓ ggplot2   3.4.3     ✓ tibble    3.2.1
## ✓ lubridate 1.9.3     ✓ tidyr     1.3.0
## ✓ purrr     1.0.2
## ── Conflicts ────────────────────────────────────────── tidyverse_conflicts() ──
## ✖ dplyr::filter() masks stats::filter()
## ✖ dplyr::lag()    masks stats::lag()
## ℹ Use the conflicted package (<http://conflicted.r-lib.org/>) to force
all conflicts to become errors
```

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

For my project I measured my daily oxygen saturation as my main metric and compared to whether liquid consumption would affect this.
* Some guidelines i followed to make it easier: I only counted drinks as liquids, did not count food including soups. I used premeasured drinks and bottles as a way to estimate my drinking.
I collected data from nov. 15 - nov. 28 of 2023
* I used an oximetry (spo2) meter to measure it daily usually at night after a day of drinking.
* My question is: Test if the data provides evidence that the average oxygen saturation level after drinking less than 16 ounces of liquid is significantly different from that of drinking 16 or more ounces.
I used the significance level of .05 to test the claim
$H_0: \mu 1 = \mu 2$ vs $H_a: \mu 1 \neq \mu 2$

```
oxygenData <- read_excel("/Users/machome/Library/Mobile
Documents/com~apple~CloudDocs/1. Data vis/Final Project/oxygen data.xlsx")
```

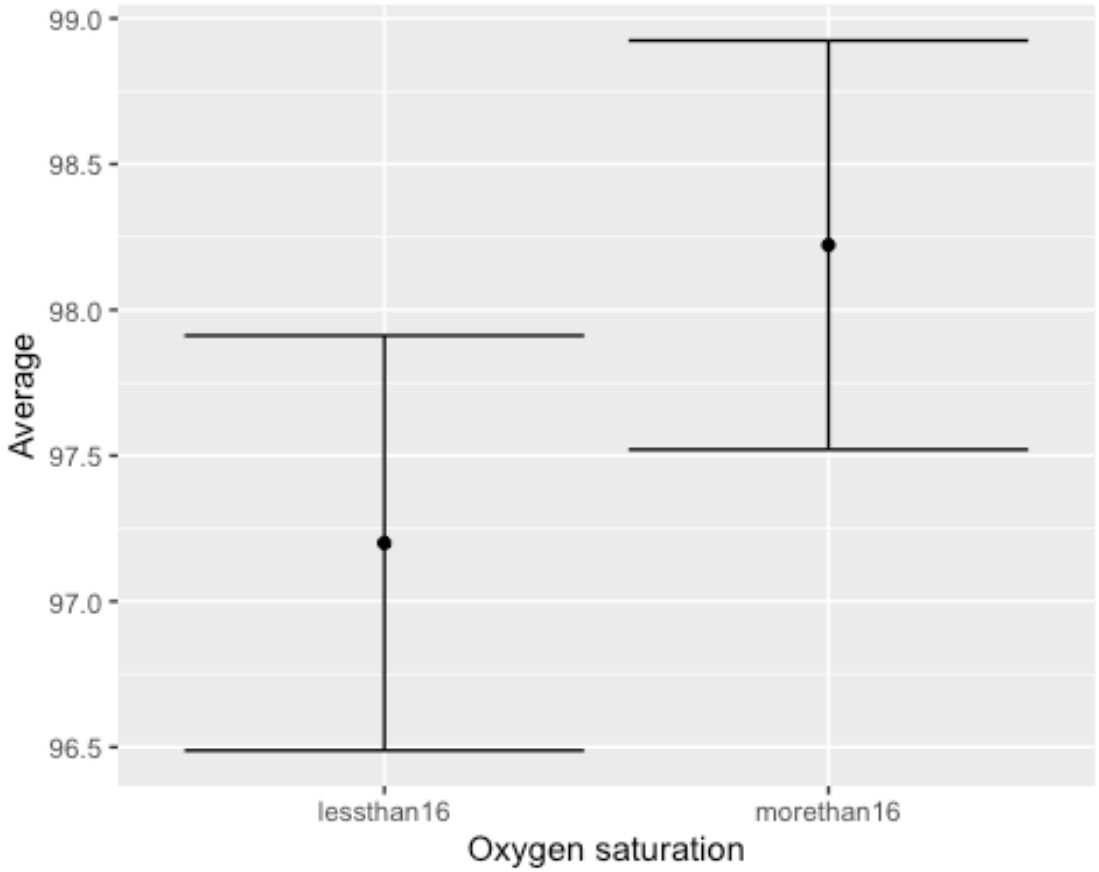
```r
t.test(OxygenSaturation~SixteenPlus, data = oxygenData, alternative =
"two.sided")
```

```
##
##  Welch Two Sample t-test
##
## data:  OxygenSaturation by SixteenPlus
## t = -4.1185, df = 8.2801, p-value = 0.003112
## alternative hypothesis: true difference in means between group lessthan16
and group morethan16 is not equal to 0
## 95 percent confidence interval:
##  -1.5912273 -0.4532171
## sample estimates:
## mean in group lessthan16 mean in group morethan16
##                 97.20000                 98.22222
```

According to two sided t-test at 5% sig. level, p value of .003112 is smaller than .05 and thus we reject H_0 in favor of the H_a and so there is significant difference in oxygen saturation level between drinking 16 or more ounces of liquid.

```r
#use length to find n
# length(oxygenData)


oxygenData %>%
  group_by(SixteenPlus) %>%
  summarize(average = mean(OxygenSaturation), SE = sd(OxygenSaturation) /
sqrt(4), tstar = qt(1-.05 / 2, 4-1)) %>%
  ggplot(aes(x = SixteenPlus, y = average)) +
  geom_point() +
  geom_errorbar(aes(ymin = average - tstar * SE, ymax = average + tstar *
SE)) +
  labs(x = "Oxygen saturation", y = "Average")
```



If the CI from one group includes the mean from the other group => no significant difference
If the CI from one group does NOT include the mean from the other group => significant difference
In this case, graph shows there is a significant difference in oxygen saturation between drinking less than 16 ounces and more as neither mean overlaps in the opposite confidence interval.

Also tested whether my mean oxygen saturation is compared to 'normal' healthy levels (95 - 100%) using 95 as my base of comparison.
$H_0: \mu = 97$ vs $H_a: \mu > 97$

```
oxygenData <- read_excel("/Users/machome/Library/Mobile
Documents/com~apple~CloudDocs/1. Data vis/Final Project/oxygen data.xlsx")


t.test(oxygenData$OxygenSaturation, mu = 95, alternative = "greater")

##
##  One Sample t-test
##
## data:  oxygenData$OxygenSaturation
```

```
## t = 16.125, df = 13, p-value = 2.821e-10
## alternative hypothesis: true mean is greater than 95
## 95 percent confidence interval:
##  97.54335      Inf
## sample estimates:
## mean of x
##  97.85714
```

According to t-test, p value is less than 5% sig. level ($p < .05$) therefore I reject H_0 in favor of H_a and accept that the claim that mean is higher than 95% oxygen saturation.