Homework2

Yuki Joyama

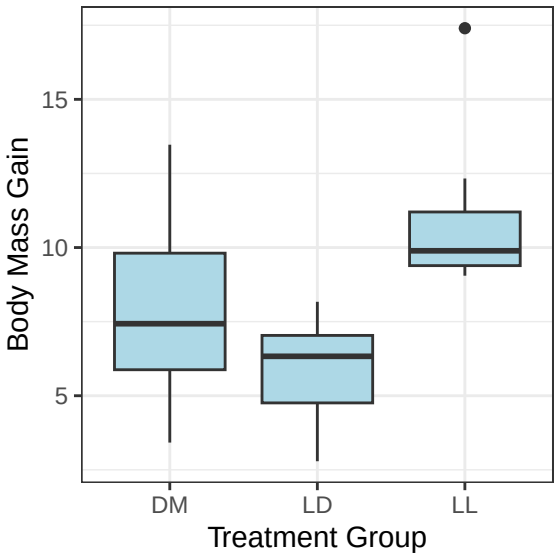
```
# libraries
library(tidyverse)
library(ggplot2)

# setup plot theme
theme_set(
  theme_bw() +
    theme(legend.position = "top")
  )

# import data
df = read_csv('./data/light.csv')
```

1. I will use boxplot to show the outcome by treatment group.

```
# boxplot
p1 = ggplot(df, aes(x = Light, y = BMGain)) +
  geom_boxplot(fill = "lightblue") +
  labs(x = "Treatment Group", y = "Body Mass Gain")
p1
```



2. Here I will subset the data to only consider LD (dark light) and LL (bright light) groups.

```r
# filter by these two groups
df2 = df |>
  filter(Light == 'LL' | Light == 'LD')
summary(df2)
```

```
##     Light               BMGain       Corticosterone       DayPct
##  Length:17          Min.   : 2.790   Min.   :  3.00   Min.   :21.85
##  Class :character   1st Qu.: 6.340   1st Qu.: 23.40   1st Qu.:40.50
##  Mode  :character   Median : 9.050   Median : 52.00   Median :61.45
##                     Mean   : 8.618   Mean   : 59.86   Mean   :57.49
##                     3rd Qu.: 9.890   3rd Qu.: 70.47   3rd Qu.:81.60
##                     Max.   :17.400   Max.   :191.22   Max.   :87.26
##   Consumption     GlucoseInt           GTT15           GTT120
##  Min.   :3.387   Length:17          Min.   :226.6   Min.   :118.3
##  1st Qu.:3.791   Class :character   1st Qu.:280.0   1st Qu.:153.7
##  Median :4.240   Mode  :character   Median :348.8   Median :227.3
##  Mean   :4.427                      Mean   :347.8   Mean   :251.8
##  3rd Qu.:4.873                      3rd Qu.:392.4   3rd Qu.:328.7
##  Max.   :7.177                      Max.   :500.0   Max.   :470.2
##     Activity
##  Min.   : 153
##  1st Qu.: 877
##  Median :1649
##  Mean   :2660
##  3rd Qu.:4482
##  Max.   :6702
```

3. I will redefine the variables using generic names as follows:

- LL group: $A = 1$
- LD group: $A = 0$
- BMGain: Y_{obs}
- Number of mice in LL group: N_1
- Number of mice in LD group: N_0

```r
# edit df2 accordingly
df2 = df2 |>
  mutate(A = ifelse(Light == "LL", 1, 0), # add variable A and input 1 for LL, 0 for LD group
         Y_obs = BMGain) |> # add new outcome column
  select(-Light, -BMGain)

n1 = sum(df2$A == 1) # number of mice in LL
n0 = sum(df2$A == 0) # number of mice in LL
```

To evaluate the causal effect of light at night on weight gain,

4.

- Mean of the outcome variable for LL group: $\bar{Y}_1^{obs}$
- Mean of the outcome variable for LD group: $\bar{Y}_0^{obs}$