P8122 HW2

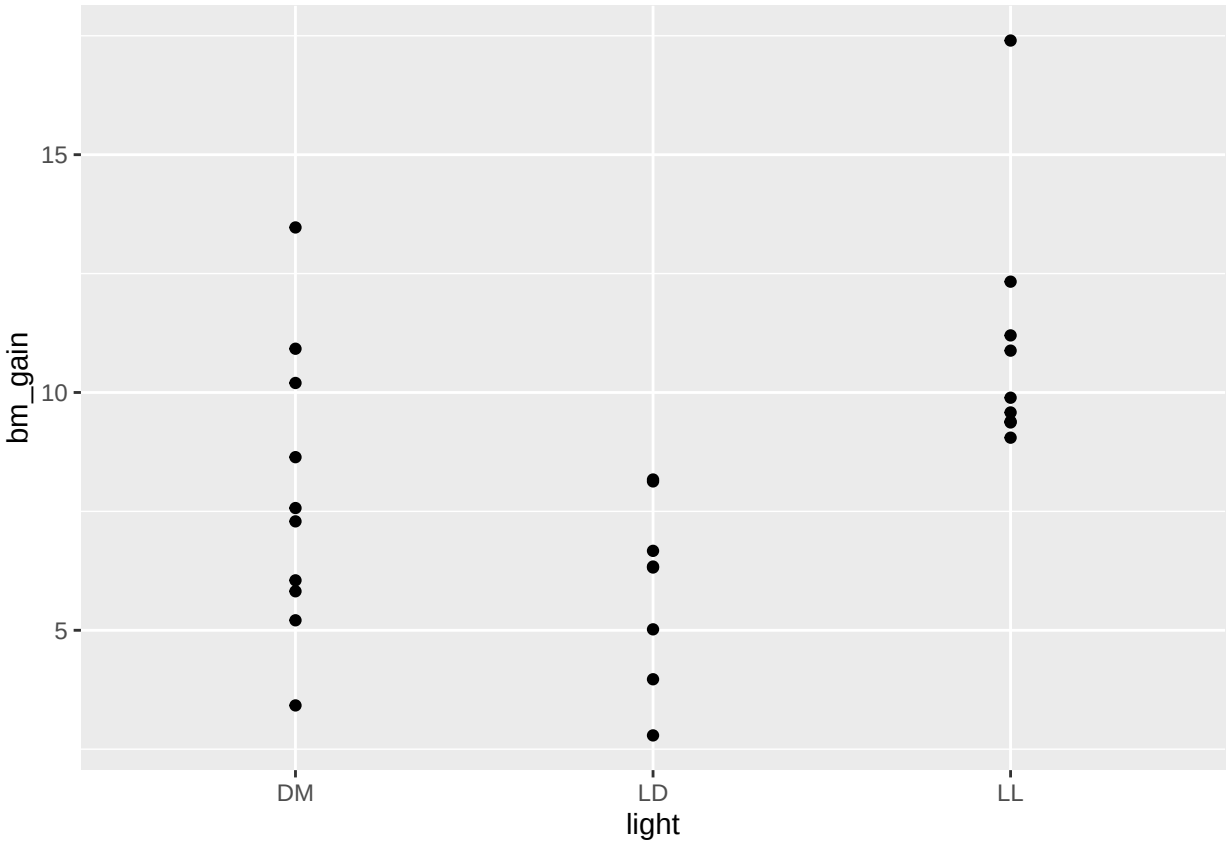
Sabrina Lin stl2137

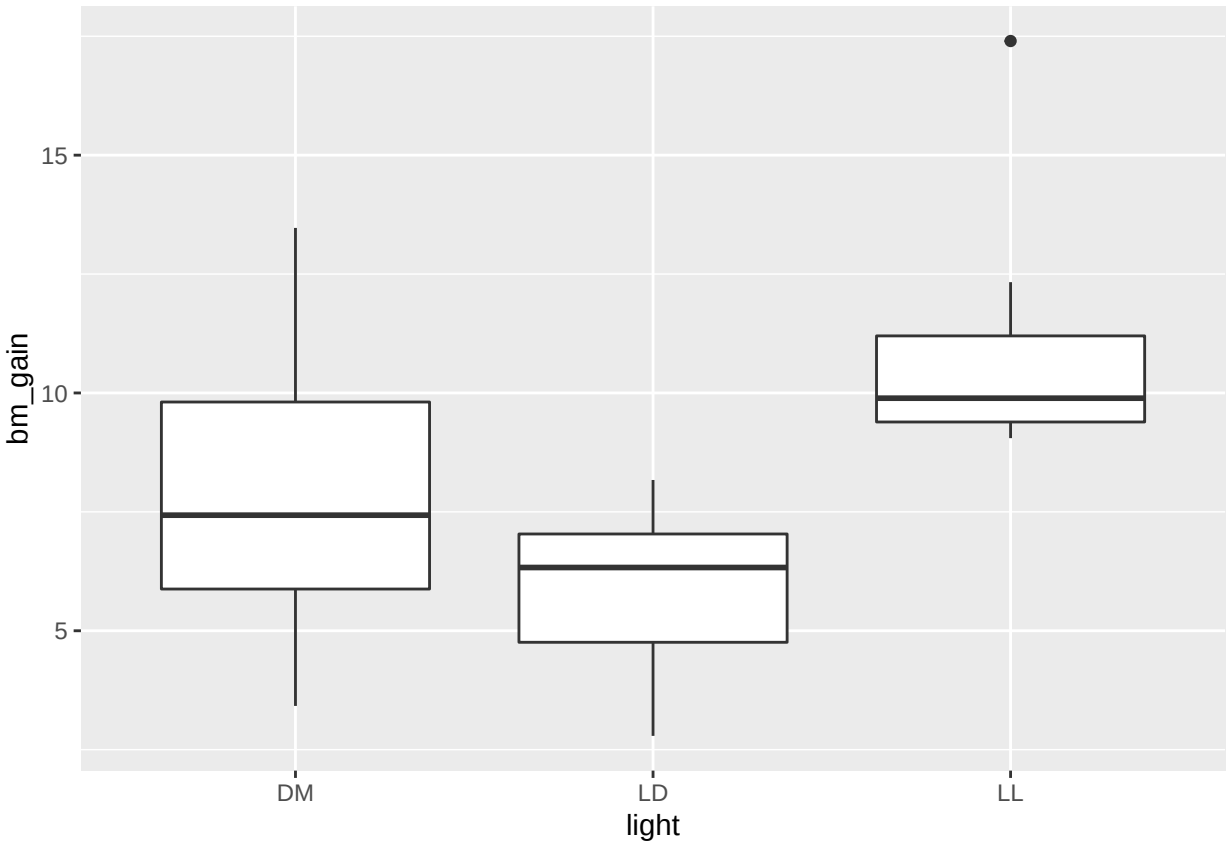
10/7/2020

Question 1

We are interested in the causal effect of light at night on weight gain. Plot the outcome by treatment group.

- DM= dark
- LD =dim
- LL =bright





Question 2

Here we will compare the mice exposed to darkness to the mice exposed to bright light overnight (once you have the code it is easy to rerun the analysis for the dim light group, if you are inter- ested). Subset the data to only consider these two groups.

Question 3

Set up the data such that everything you will need has generic names (such as Yobs or whatever you want to call them). Everything specific to the context of your data (variable names, sample sizes) should only be in your R Script here. Everything else should be generic so you can copy/paste it for later use. What quantities will you need to evaluate the causal effect of light at night on weight gain?

- `light` is the treatment assignment variable and `bm_gain` is the outcome variable.

Question 4

Suppose we want the statistic to be the difference in means between the two treatment groups. Calculate Tobs.

- The difference in means between the 2 treatment groups, T_{obs}, is -3.151.

Question 5

How many different possibilities are there for A? Enumerate all of these possibilities in a matrix. (Hint: it's probably easiest to first install the ri or perm package, have a look at the function chooseMatrix in R, it may come in handy.)

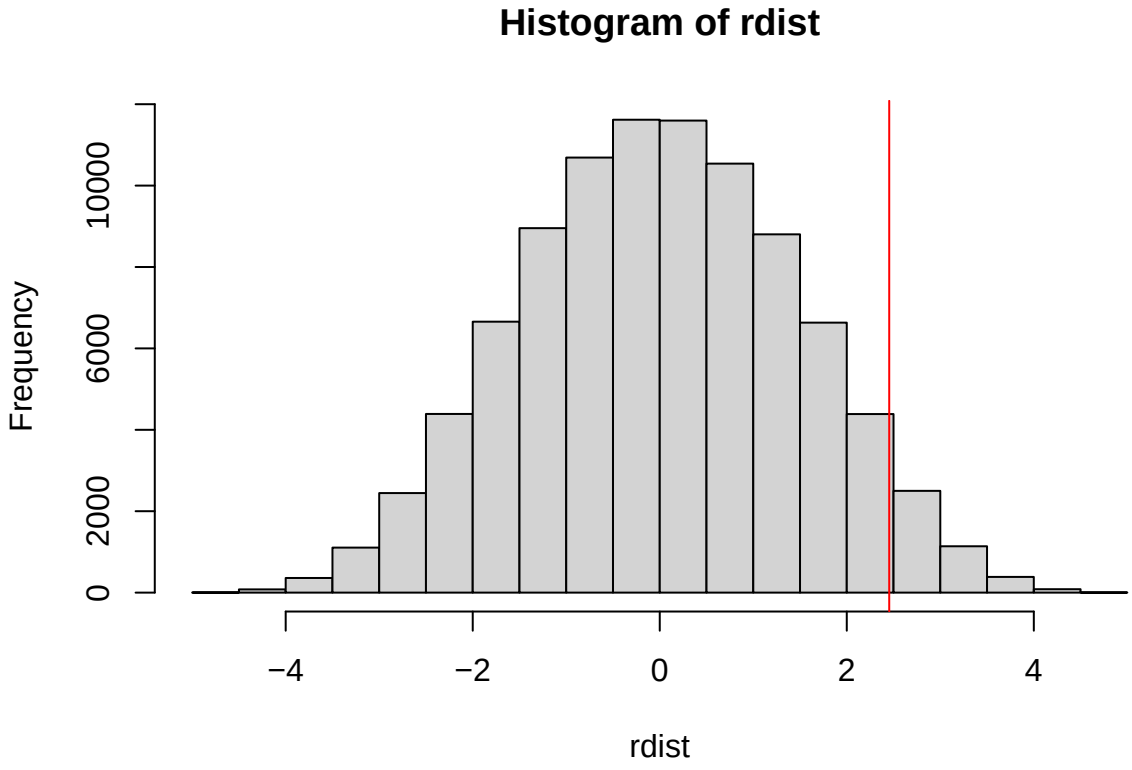
Question 6

State the sharp null hypothesis of no difference. Calculate the test statistic under one of these possibilities for A (the first one), under the sharp null hypothesis.

- The sharp null hypothesis of no difference is that there is no treatment effect between Y_0 and Y_1 for each individual observation.
- Mathematical representation: $H_0 : \tau_i = Y_{0i} - Y_{1i} = 0$

Question 7 & 8

Generate the exact randomization distribution for T, under the sharp null hypothesis of no difference. Plot this distribution, and mark the observed test statistic.



Question 9

Calculate the exact p-value, based on this distribution.

The exact p-value is 0.0478685.

Question 10

What do you conclude?

- Because the null hypothesis is rejected, I conclude that there is a significant treatment difference between dim light and bright light in regards to weight gain for the mice.

Appendix

```
knitr::opts_chunk$set(echo = FALSE)
library(tidyverse)
library(ri)
### Import data
mice_dat <- read_csv("/Users/SabrinaLin/Documents/Fall_2020_Causal_Inference/Homework/HW2/light.csv") %
### Scatterplot
mice_dat %>%
  ggplot(aes(x = light, y = bm_gain)) +
  geom_point()

### Boxplot
mice_dat %>%
  ggplot(aes(x = light, y = bm_gain)) +
  geom_boxplot()
### Subsetting data to only contain dark & bright light exposured mice
dark_bright_mice <- mice_dat %>%
  filter(
    light %in% c("DM", "LL")
  )
### y_obs: bm_gain; A_treat: light;

dat <- dark_bright_mice %>%
  rename(
    y_obs = bm_gain,
    a_treat = light
  ) %>%
  select(a_treat, y_obs) %>%
  arrange(a_treat)

Y <- dat$y_obs
a_treat <- dat$a_treat
group_mean_dat <- aggregate(Y, list(a_treat), mean)

t_obs <- group_mean_dat[1, 2] - group_mean_dat[2, 2]
# Assigning 0 = DM, 1 = LL
a <- c(rep(0, 10), rep(1, 9))
a_bold <- genperms(a, maxiter = 92378) # choose(19, 10) = 92378

group_mean_sharp <- aggregate(Y, list(a_bold[,1]), mean)

t_obs_sharp <- group_mean_sharp[1, 2] - group_mean_sharp[2, 2]
rdist <- rep(NA, times = ncol(a_bold))
```

```r
for (i in 1:ncol(a_bold)){
  a_tilde <- a_bold[,i]
  rdist[i] <- mean(Y[a_tilde == 1]) - mean(Y[a_tilde == 0])
}

hist(rdist)
abline(v = t_obs_sharp, col = "red")
pval <- mean(rdist >= t_obs_sharp)
```