P8122 Homework-2 yz4184

Yunlin Zhou

2022-09-30

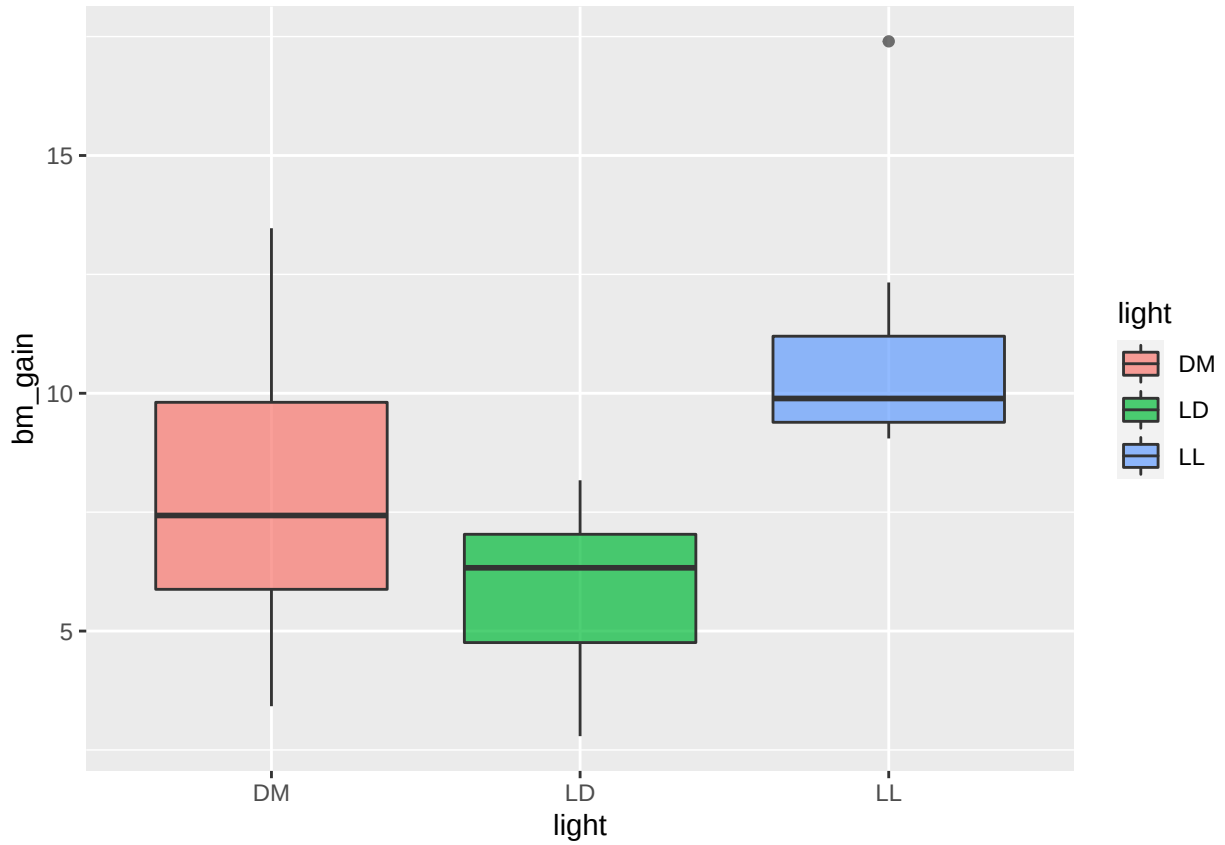
```
data = read.csv("./light.csv") %>%
  janitor::clean_names()
```

1. (5 points) We are interested in the causal effect of light at night on weight gain. Plot the outcome by treatment group.

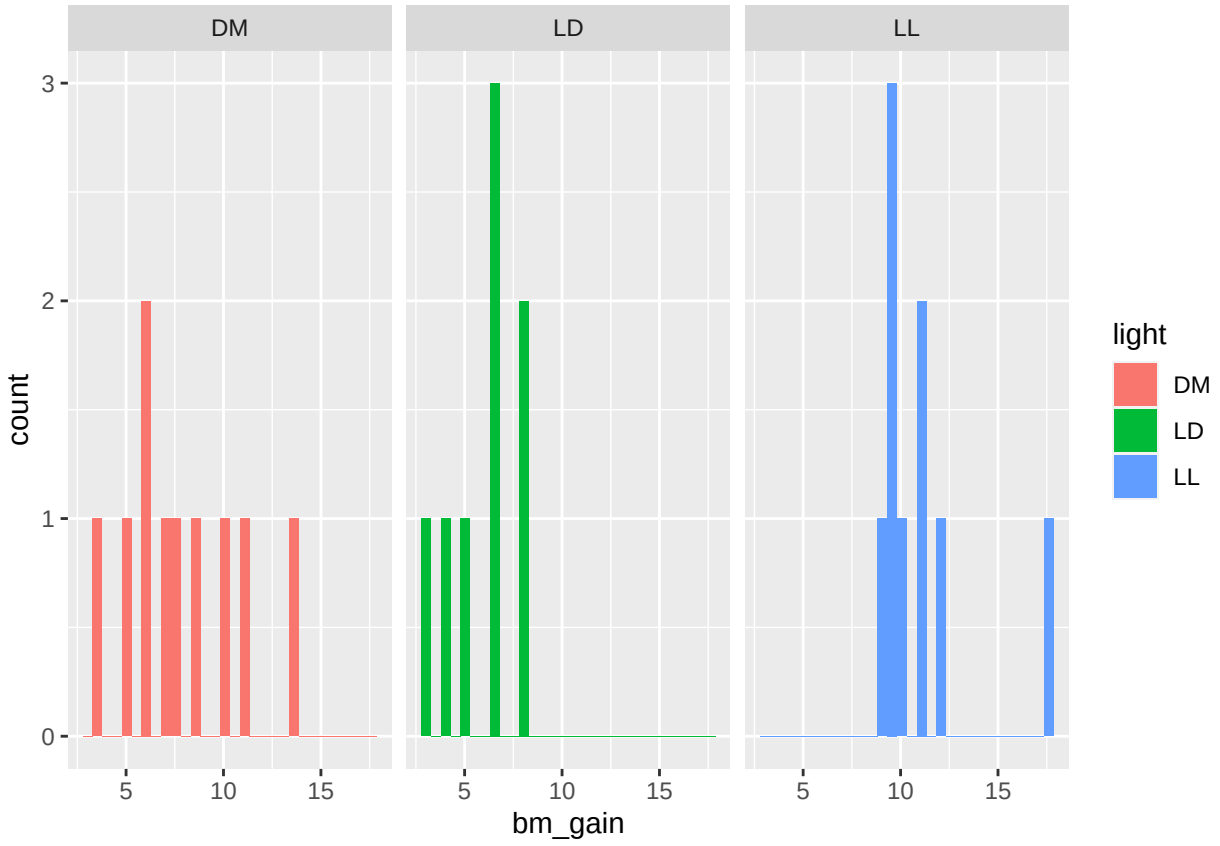
- DM= dim
- LD =dark
- LL =bright

From the plots below we can conclude that, the bright group is tending to gain the most weight; the dark group is tending to gain the least weight.

```
data %>%
  ggplot(aes(x = light, y = bm_gain, fill = light)) +
  geom_boxplot(alpha = 0.7)
```




```
data %>%
  ggplot(aes(x = bm_gain, fill = light)) + geom_histogram()+facet_grid(. ~ light)
```



2. (5 points) 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 interested). Subset the data to only consider these two groups.

```
dark_df = data %>%
  filter(light %in% c("LD"))

bright_df = data %>%
  filter(light %in% c("LL"))

dim_df = data %>%
  filter(light %in% c("DM"))

bright_dark_df = data %>%
  filter(light %in% c("LD", "LL"))
```

3. (15 points) 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?

```
bright_dark_df = bright_dark_df  %>%
  select(bm_gain, light, everything())%>%
  rename( Y_obs = bm_gain,
          A = light)
```

```
Y_obs = bright_dark_df$Y_obs
A = bright_dark_df$A
A = as.factor(A)
```

4. (10 points) Suppose we want the statistic to be the difference in means between the two treatment groups. Calculate Tobs.

```
T_obs = mean(Y_obs[A == "LL"]) - mean(Y_obs[A == "LD"])
T_obs
```

```
## [1] 5.08375
```

The difference in means between the 2 treatment groups is 5.08375.

5. (10 points) How many different possibilities are there for A? Enumerate all of these possibilities ina matrix. (Hint: it's probably easiest to first install the ri or perm package, have a look at the function choose Matrix in R, it may come in handy.)

```
Abold = chooseMatrix(17, 9)
Abold = t(Abold)
ncol(Abold)
```

```
## [1] 24310
```

There are 24310 possibilities for A.

6. (15 points) 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.

Sharp null hypothesis: For each individual observation, there is no treatment effect.

H0:Ti = Yi1-Yi0

```
A_tilde_1 <- Abold[, 1]
T_sharp <- mean(Y_obs[A_tilde_1 == 1]) - mean(Y_obs[A_tilde_1 == 0])
T_sharp
```

```
## [1] 1.551528
```

The t-statistic under the sharp null hypothesis is 1.5515278.

7. (10 points) Generate the exact randomization distribution for T, under the sharp null hypothesis of no difference.

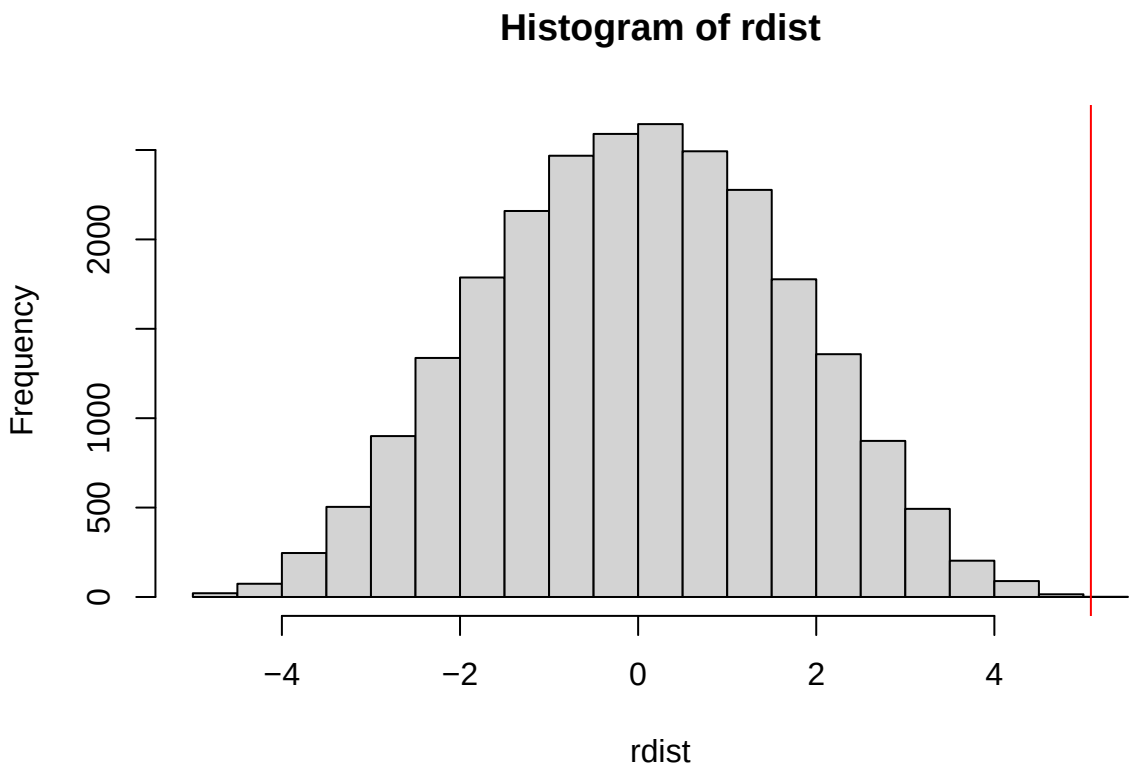
```
rdist <- rep(NA, times = ncol(Abold))
for (i in 1:ncol(Abold)) {
  A_tilde <- Abold[, i]
  rdist[i] <- mean(Y_obs[A_tilde == 1]) - mean(Y_obs[A_tilde == 0])
}

head(rdist)
```

```
## [1]  1.551527778 -0.700972222  0.290694444 -0.006805556 -0.358611111
## [6]  0.713333333
```

8. (10 points) Plot this distribution, and mark the observed test statistic.

```
hist(rdist)
abline(v = T_obs,col="red")
```



9. (10 points) Calculate the exact p-value, based on this distribution.

```
pval <- mean(rdist >= T_obs)
pval
```

```
## [1] 4.113534e-05
```

The exact p-value based on this distribution is 4.1135335×10^{-5}.

10. (10 points) What do you conclude?

The p-value is 4.1135335×10^{-5}, which is less than 0.05. So we rejected the null hypothesis that for each individual observation, there is no treatment effect. And we conclude that there is sufficient evidence to say that the light at night played a causal role in the obesity epidemic.