

hw2

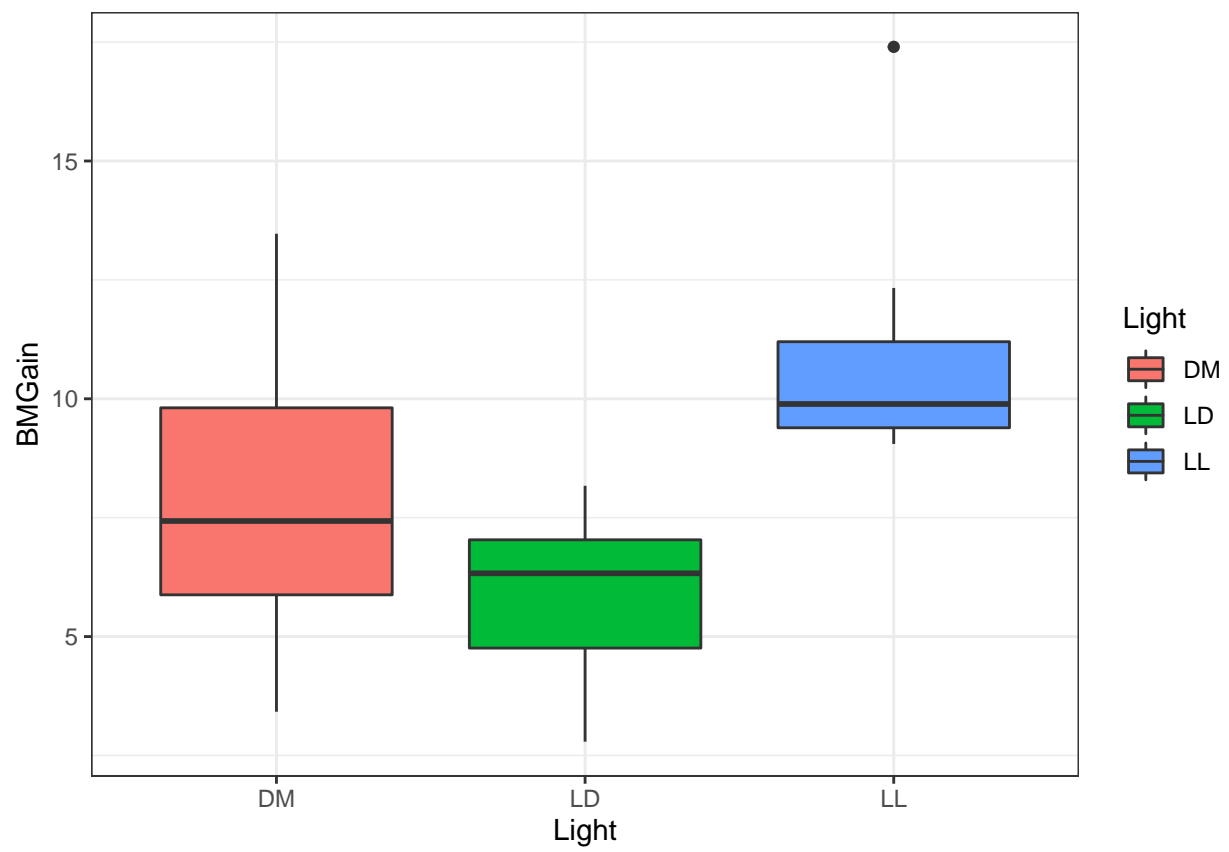
ZiqianHe

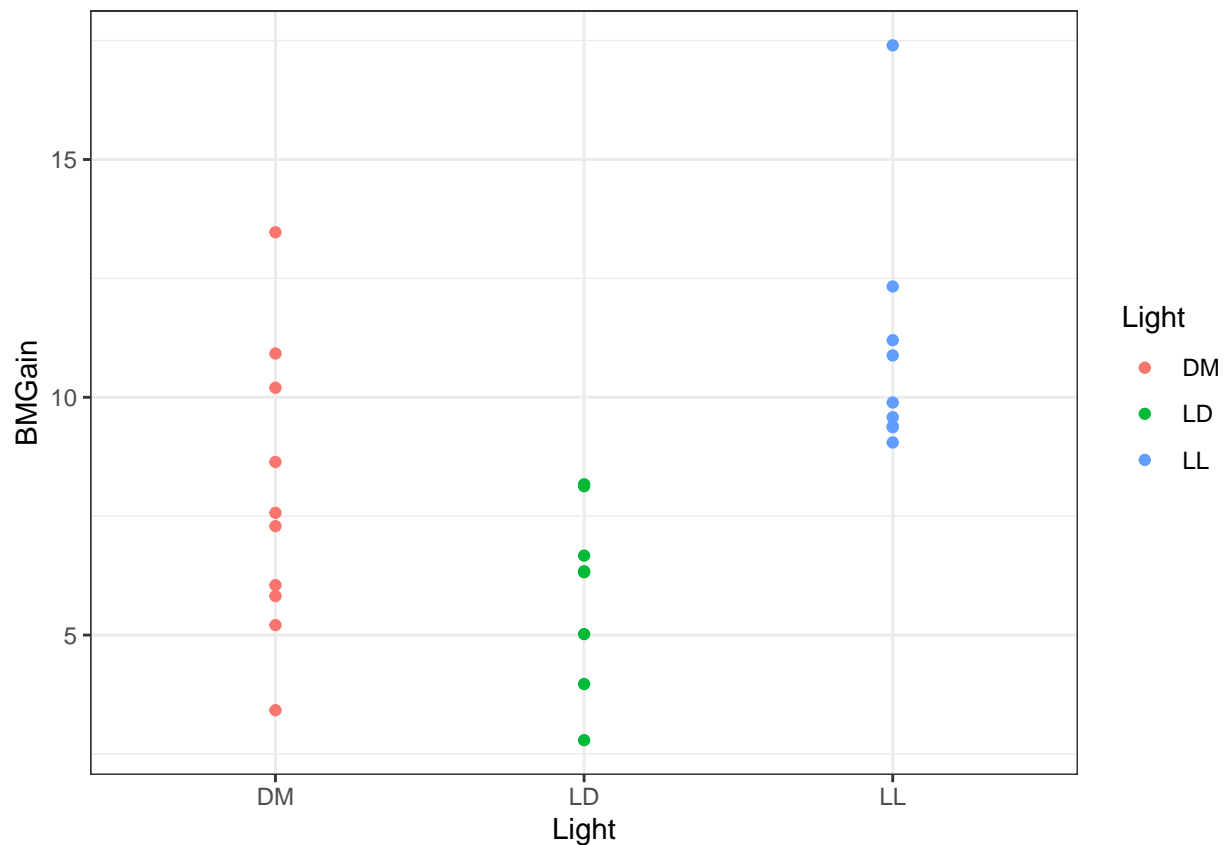
1

Plot the outcome by treatment group

*outcome:BMGain

*treatment group:Light(DM=dim, LD=dark, LL=bright)





*Different groups have different outcome. The median body mass gain in the LL group is the highest, and LD group is the lowest. The median body mass gain in the DM and LL are similar but overall LD group is lower than DM group.

2

compare the mice exposed to darkness to the mice exposed to bright light overnight. Subset the data to only consider these two groups.

Light	BMGain	Corticosterone	DayPct	Consumption	GlucoseInt	GTT15	GTT120	Activity
LD	5.02	87.838	31.063	3.791	No	228.448	134.483	1437
LD	6.67	191.220	41.408	3.923	No	231.183	220.430	2541
LD	8.17	67.700	47.573	4.489	No	226.563	141.406	346
LD	2.79	41.017	34.947	4.161	No	323.077	199.038	5837
LD	8.13	21.817	41.940	4.416	No	500.000	190.361	877
LD	6.34	23.403	40.500	4.890	No	280.000	118.333	1649
LD	6.32	70.470	28.950	4.946	No	299.174	153.719	728
LD	3.97	56.718	21.846	4.004	No	461.250	230.000	6048
LL	9.89	42.132	71.552	3.387	Yes	378.704	328.704	5752
LL	9.58	48.238	61.453	3.451	No	379.091	227.273	1256
LL	11.20	92.191	85.978	3.501	Yes	366.129	383.871	244
LL	9.05	51.999	64.827	4.240	No	392.373	250.000	931
LL	12.33	12.252	81.600	3.479	Yes	466.346	470.192	3582
LL	9.39	3.000	87.257	5.940	Yes	259.615	413.462	2657

Light	BMGain	Corticosterone	DayPct	Consumption	GlucoseInt	GTT15	GTT120	Activity
LL	10.88	132.400	70.441	4.586	No	348.780	126.016	153
LL	9.37	8.615	84.415	4.873	Yes	335.652	286.957	4482
LL	17.40	66.679	81.636	7.177	Yes	435.644	405.941	6702

*In total, 17 observations. 9 in the bright light group and 8 in the darkness group.

3

set up the data such that everything you will need has generic names (such as Yobs or whatever you want to call them).

```
dm_ll_generic<-dm_ll %>% rename(Y_obs= BMGain, A = Light) %>% arrange(A) %>% mutate(A=ifelse(A=="LL",1,0))
Y_obs = dm_ll_generic$Y_obs
A=dm_ll_generic$A
```

*Y_obsThe is outcome of BMGain and A is the treatment assignment of Light (light = 1 and dark = 0).

4

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

*The difference in means between the Two treatment groups, T_{obs} , is 5.08375.

5

How many different possibilities are there for A? Enumerate all of these possibilities in a matrix.

```
##      [,1] [,2] [,3] [,4] [,5]
## [1,]    0    0    0    0    0
## [2,]    0    0    0    0    0
## [3,]    0    0    0    0    0
## [4,]    0    0    0    0    0
## [5,]    0    0    0    0    0
## [6,]    0    0    0    0    0
## [7,]    0    0    0    0    0
## [8,]    0    1    1    1    1
## [9,]    1    0    1    1    1
## [10,]   1    1    0    1    1
## [11,]   1    1    1    0    1
## [12,]   1    1    1    1    0
## [13,]   1    1    1    1    1
## [14,]   1    1    1    1    1
## [15,]   1    1    1    1    1
## [16,]   1    1    1    1    1
## [17,]   1    1    1    1    1
```

*In total there are 17 subjects so we have 24310 and above is 5 of them.

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.

*Sharp null hypothesis: there is no treatment effect, so the vector of observed outcomes Y dose not change with different A. $H_0 : \tau_i = Y_{0i} - Y_{1i} = 0$

*The test statistic under one of these probabilities for A (the first one), under the sharp null hypothesis is 5.08375

7

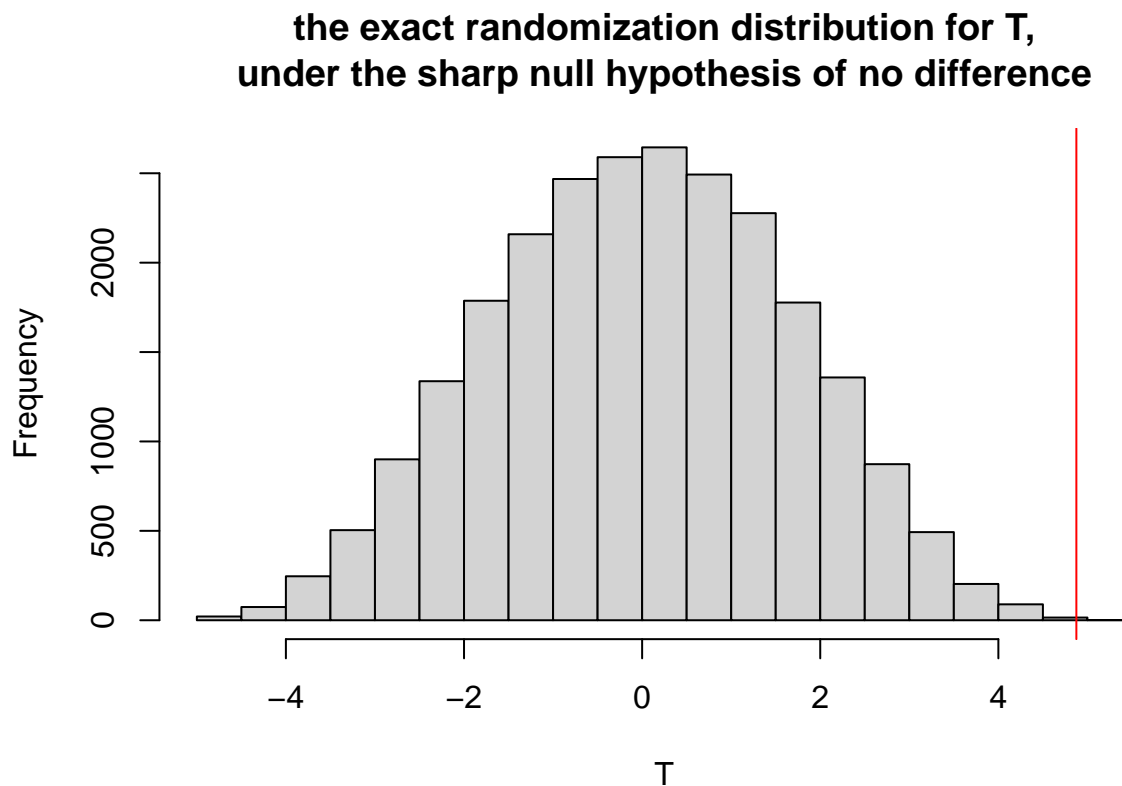
Generate the exact randomization distribution for T, under the sharp null hypothesis of no difference.

```
## [1] 5.083750 3.685972 3.759167 3.376667 3.884306 3.109861 3.804028 3.452222
## [9] 3.808750 1.912778 4.240833 4.314028 3.931528 4.439167 3.664722 4.358889
## [17] 4.007083 4.363611 2.467639 2.916250
```

*Above are the first 20 values of T, under the sharp null hypothesis.

8

Plot this distribution, and mark the observed test statistic.



9

Calculate the exact p-value, based on this distribution

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

10

What do you conclude?

- Our observed test statistic is extreme in the exact randomization distribution for T. The p-value < 0.05 , so we can reject the sharp null hypothesis and conclude that there is an individual effect of treatment group of bright light VS darkness on body mass gain in mice.

code

```
#load data
data=fread("D:/columbia/term3/casual/P8122/HW2/light.csv")

#1
ggplot(data,aes(x=Light,y=BMGain,fill=Light))+geom_boxplot()+theme_bw()
ggplot(data,aes(x=Light,y=BMGain,color=Light))+geom_point()+theme_bw()

#2
dm_ll <- data[Light %in% c("DM","LL")]
dm_ll %>% arrange(Light) %>% knitr::kable()

#3
dm_ll_generic<-dm_ll %>% rename(Y_obs= BMGain, A = Light) %>%arrange(A) %>% mutate(A=ifelse(A=="LL",1,0))
Y_obs = dm_ll_generic$Y_obs
A=dm_ll_generic$A

#4
T_obs=mean(Y_obs[A==1])-mean(Y_obs[A==0])

#5
A_poss = chooseMatrix(19,9)
A_poss = t(A_poss)
A_poss[,1:5]

#6
A_6=A_poss[,1]
T_6=mean(Y_obs[A_6==1])-mean(Y_obs[A_6==0])

#7
rdist <- rep(NA, times = ncol(A_poss))
#run a for loop through each randomization scenario and calculate the corresponding test statistic, then
for (i in 1:ncol(A_poss)) {
  A_tilde <- A_poss[, i]
  rdist[i] <- mean(Y_obs[A_tilde == 1]) - mean(Y_obs[A_tilde == 0])
}
```

```

}
rdist[1:20]

#8
pval <- mean(rdist >= T_obs)
quant <- quantile(rdist,probs = 1-pval)
hist(rdist,xlab="T",main="the exact randomization distribution for T,\nunder the sharp null hypothesis")
abline(v = quant,col="red")

#9
pval <- mean(rdist >= T_obs)
pval

```