HW6 - MattMaslow

Matthew Maslow

3/28/2023

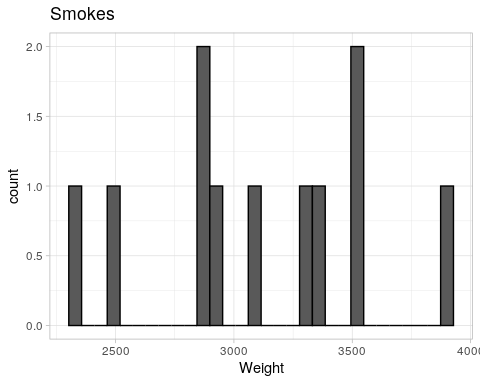
## 7.6

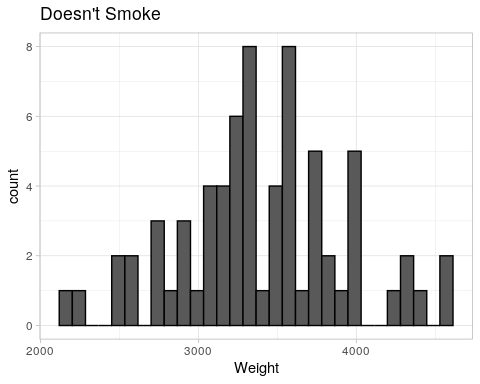
#### 12.

###### a.

data("Girls2004")

smoke <-   
 Girls2004 %>% filter(  
 Smoker == "Yes"  
 ) %>%   
 ggplot(smoke, mapping = aes(Weight) ) +  
 geom\_histogram(col="black") +  
 labs(title = "Smokes") + theme\_light()  
noSmoke <- Girls2004 %>% filter(  
 Smoker == "No"  
 ) %>%   
 ggplot(noSmoke, mapping = aes(Weight) ) +  
 geom\_histogram(col = "black") +  
 labs(title = "Doesn't Smoke") + theme\_light()





###### b.

# 95% CI for lower bound of the diff in means of baby weights of  
# smokes vs. non-smokers  
NSbar = mean(noSmoke$data$Weight)  
Smkbar = mean(smoke$data$Weight)  
sNS = sd(noSmoke$data$Weight)  
sSmk = sd(smoke$data$Weight)  
(df = ( (sNS^2/ 69)^2 + (sSmk^2 / 11)^2 ) / ( ( ( (sNS^2/ 69)^2) / 68 + (sSmk^2 / 11)^2 / 10) ) )

## [1] 10.3442

t = -2.218  
result = (NSbar - Smkbar) + (t) \* (sqrt( (sNS^2/69) + (sSmk^2/11) ) )  
result

## [1] -56.11419

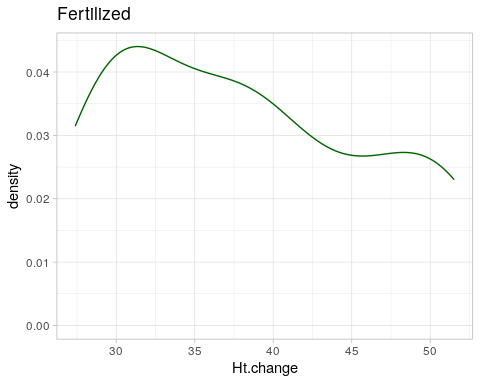
We are 95% confident the that the lower bound of the mean difference of baby weights of non-smokers versus smokers is -56.114 grams.

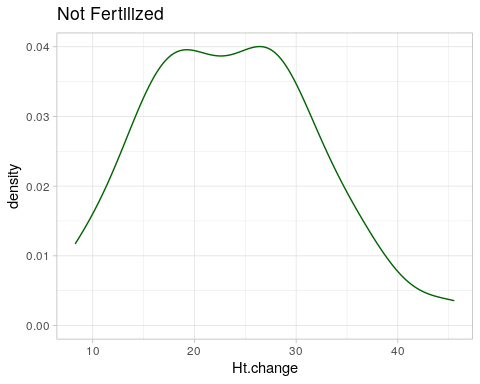
#### 13.

data(Spruce)

###### a.

fertilized <-   
 Spruce %>% filter(  
 Fertilizer == "F"  
 ) %>%   
 ggplot(fertilized, mapping = aes(Ht.change) ) +  
 geom\_density(col = "darkgreen")+  
 labs(title = "Fertilized" ) + theme\_light()  
nonFertilized <-   
 Spruce %>% filter(  
 Fertilizer == "NF"  
 ) %>%   
 ggplot(nonFertilized, mapping = aes(Ht.change) ) +  
 geom\_density(col = "darkgreen")+  
 labs(title = "Not Fertilized") + theme\_light()





###### b.

# 95% CI for the lower bound of the diff in mean height change for the Spruces  
FertBar <- mean(fertilized$data$Ht.change)  
nonFertBar <- mean(nonFertilized$data$Ht.change)  
sFert <- sd(fertilized$data$Ht.change)  
sNonFert <- sd(nonFertilized$data$Ht.change)  
(df = ( (sFert^2/ 36)^2 + (sNonFert^2 / 36)^2 ) / ( ( ( (sFert^2/ 36)^2) / 35 + (sNonFert^2 / 36)^2 / 35) ) )

## [1] 35

t = -2.030  
answer = (FertBar - nonFertBar) + (t) \* (sqrt( (sFert^2/36) + (sNonFert^2/36) ) )  
answer

## [1] 10.76017

We are 95% confident that the lower bound of the mean difference between Spruces with fertilizer versus without is 10.76 feet.

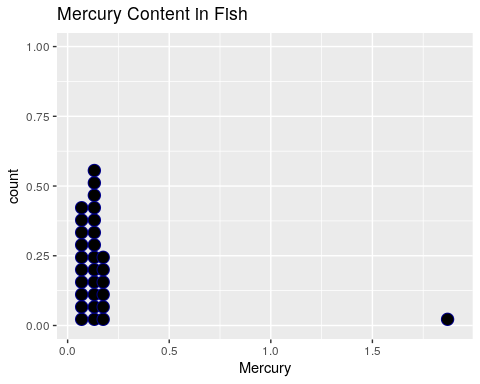
## 5.6:

#### 12.

data("FishMercury")

###### a.

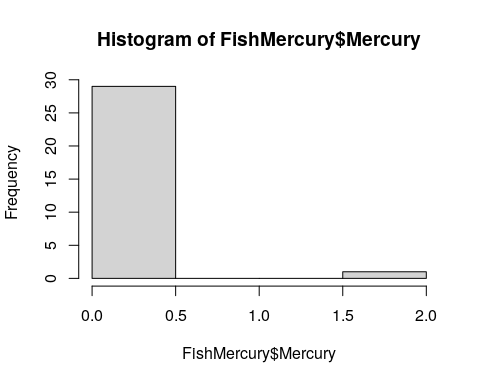
fmPlot <-  
 FishMercury %>%  
 ggplot(aes(Mercury)) +  
 geom\_dotplot(col = "darkblue") +  
 labs(title = "Mercury Content in Fish")



Most of the data on the fishes are located under 0.25mg, then there is an outlier that is above 1.75mg.

###### b.

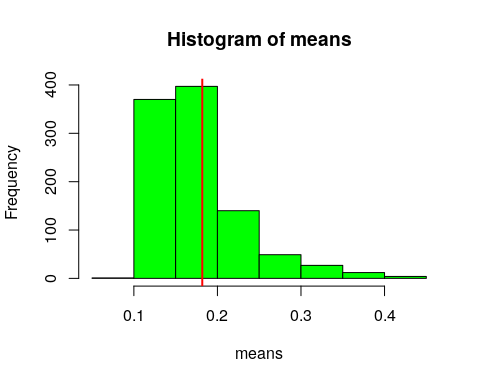
hist(FishMercury$Mercury)



mean(FishMercury$Mercury)

## [1] 0.1818667

n=length(FishMercury$Mercury)  
means = NULL  
B = 1000  
for (i in 1:B)  
{  
 boot\_sample = sample(FishMercury$Mercury,n,replace=T)  
 boot\_mean = mean(boot\_sample)  
 means = c(means, boot\_mean)  
}  
hist(means,col="green")  
abline(v=mean(FishMercury$Mercury),col="red",lwd=2)



# mean of bootstrap distribution (should be close to our statistic)  
mean(means)

## [1] 0.1797776

sd(boot\_sample)

## [1] 0.3221254

# 95% CI  
quantile(means,c(0.025, 0.975))

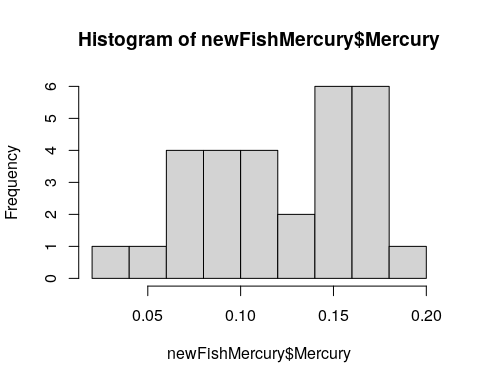
## 2.5% 97.5%   
## 0.1126242 0.3096250

Bootstrap percentile interval for 95% is

###### c.

newFishMercury <-  
 FishMercury %>%  
 filter(Mercury < 1)

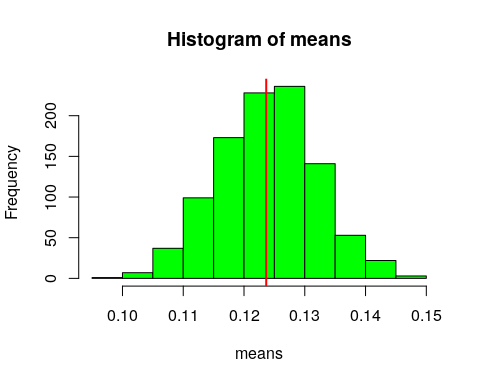
hist(newFishMercury$Mercury)



mean(newFishMercury$Mercury)

## [1] 0.1236552

n=length(newFishMercury$Mercury)  
means = NULL  
B = 1000  
for (i in 1:B)  
{  
 boot\_sample = sample(newFishMercury$Mercury,n,replace=T)  
 boot\_mean = mean(boot\_sample)  
 means = c(means, boot\_mean)  
}  
  
hist(means,col="green")  
abline(v=mean(newFishMercury$Mercury),col="red",lwd=2)



# mean of bootstrap distribution (should be close to our statistic)  
mean(means)

## [1] 0.1238498

sd(boot\_sample)

## [1] 0.03885314

# 95% CI  
quantile(means,c(0.025, 0.975))

## 2.5% 97.5%   
## 0.107994 0.139969

Bootstrap percentile interval for 95% without the outlier is…

###### d.

By removing the outlier it significantly improved the distribution of data. The mean of the data is now centered in the histogram, which also has a normal shape now, where the original was skewed. Also, looking at 95% bootstrap interval it narrowed the width of bounds.