Lab 3

Nathaniel Tomasetti

8 August 2016

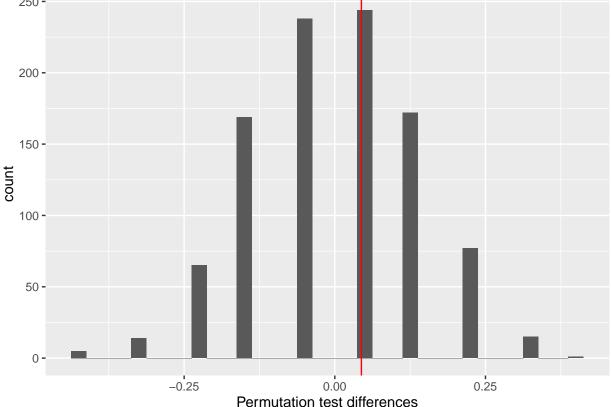
group	no	yes	total
control	12	4	16
treatment	24	10	34

```
yawn_expt %>%
group_by(group, yawn) %>%
tally() %>%
ungroup() %>%
spread(yawn, n, fill=0) %>%
mutate(total = rowSums(.[-1])) %>%
#Use mutate here to calculate proportions %>%
kable()
#You might want to save the table
#Do mutate(...) -> give.it.a.name before the kable()
#Then you can get at the proportions with name$proportion[1] and [2]
```

```
permutation <- function(data){
    data %>%

    #We need a line here to resample the yawn variable
    #it might look like mutate(yawn = something that resamples yawn) %>%
    group_by(group, yawn) %>%
    tally() %>%
    ungroup() %>%
    spread(yawn, n, fill=0) %>%
    mutate(total = rowSums(.[-1]), p = yes/total) %>%
    #Second mutate to get proportions
    #We need to add something to calculate the difference
    return() #Put the difference between these brackets
```

```
#So we start with an empty object
prop.difference <- NULL</pre>
for(i in 1:1000){ #You want more, but my laptop is slow.
  #Each step attaches the result of the permutation to the end
  prop.difference <- c(prop.difference, permutation(yawn_expt))</pre>
}
#another way to write the loop
prop.difference <- vector(length = 1000)</pre>
for(i in 1:1000){
  prop.difference[i] <- permutation(yawn_expt)</pre>
library(ggplot2)
#ggplot2 can do a lot of things, but we need to make a dataframe
prop.diff <- data.frame(prop.difference)</pre>
#prop.diff is the name of the dataframe, it has one variable called prop.difference
#I've saved the difference in proportions in q1 to expt.diff
ggplot(data=prop.diff) + geom_histogram(aes(x=prop.difference), binwidth = 0.025) +
  geom_vline(xintercept = expt.diff, colour = "red") + labs(x="Permutation test differences")
  250 -
  200 -
```



#1 if greater than the experimental difference, 0 otherwise prop.difference>expt.diff

```
#So add all the 1's to get the total sum(prop.difference>expt.diff)
```

[1] 265

#Find how many trials we ran - Or just put in the length of the loop length(prop.difference)

[1] 1000

#So the proportion greater than the experimental difference is
sum(prop.difference>expt.diff) / length(prop.difference)

[1] 0.265

#Lets colour in the part of the histogram that is above the experimental difference
#We add a second red histogram on top, that just has the part above expt.diff
above = filter(prop.diff, prop.difference>expt.diff)
ggplot() + geom_histogram(data = prop.diff, aes(x = prop.difference), binwidth = 0.025) +
 geom_histogram(data = above, aes(x = prop.difference), fill = "red", binwidth = 0.025) +
 geom_vline(xintercept = expt.diff, colour = "red")

