Lab6

Kristina Arevalo

Fri Oct 09 2020

Contents

Problem 1

Write a function that conducts a randomization test for the mean difference between two groups, and show that it works. Specifically, using your function, conduct a randomization test on the same data we used in the above example from lab. Report the results and briefly discuss what the results of the randomization tell you. (6 points). Extra: if the observed mean difference in the experiment was found to be .5, what would you have concluded from the randomization test?

Inputs:

the inputs should include a vector for group 1, and vector for group 2, and the number of permutations/re-samplings of the data to create.

Outputs:

output each group mean, and the difference between group means output a histogram of the sampling distribution of the possible mean differences produced by the randomization process output the probability or odds of obtaining the observed mean difference or larger.

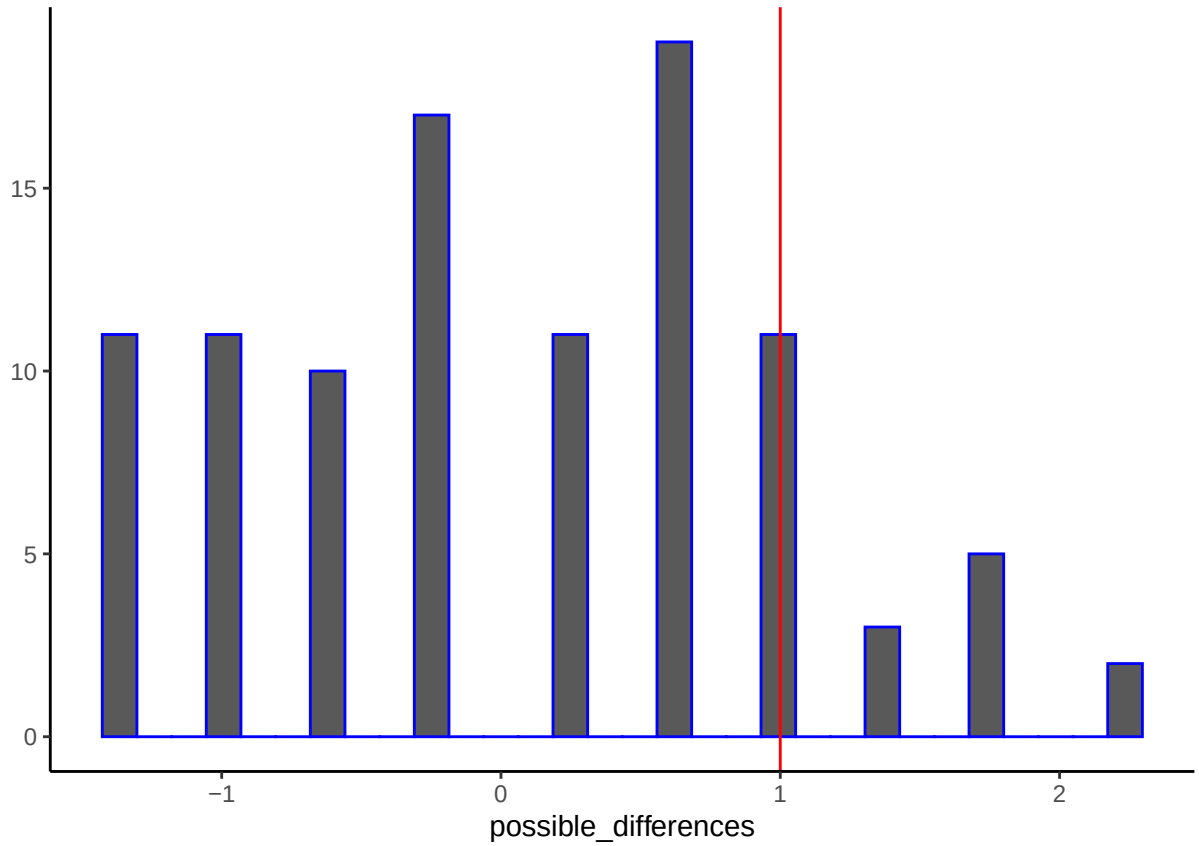
Optional:

include the ability to calculate the probability of obtaining any mean difference or larger deal with negative difference scores appropriately add one or two-tailed test options

```
group_one<- c(3,4,5,6,7)
group_two<- c(5,3,2,4,6)

randomization_test <- function(A,B,resamples=100){
  group_A_mean <- mean(A)
  group_B_mean <- mean(B)
  mean_difference <- group_A_mean - group_B_mean

  all_means <- c(A,B)

  possible_differences <- c()
for(i in 1:resamples){
  resample <- sample(all_means)
  new_A_mean <- mean(resample[1:length(A)])
  new_B_mean <- mean(resample[(length(A)+1):length(all_means)])
  possible_differences[i] <- new_A_mean-new_B_mean
}
  the_plot <- qplot(possible_differences)+
  geom_histogram(color="blue")+
  geom_vline(xintercept=mean_difference, color ="red")+
  theme_classic()

  p_value <- length(possible_differences[possible_differences >= mean_difference])/resamples

  test_output <- list(A_mean = group_A_mean,
                      B_mean = group_B_mean,
                      difference = mean_difference,
                      randomization=possible_differences,
                      plot = the_plot,
                      p_value = p_value)
  return(test_output)
}
randomization_test(A = group_one,
```

```
                  B = group_two,
                  resample=100)
```

```
## $A_mean
## [1] 5
##
## $B_mean
## [1] 4
##
## $difference
## [1] 1
##
## $randomization
##   [1]  1.8 -1.4 -1.4  1.8  0.6 -0.6 -0.2  0.2  0.2  0.2  0.6 -0.2  0.6 -1.4 -0.2
##  [16]  0.6  0.6 -0.2  0.6 -1.4 -1.4  0.2 -1.4 -1.4  0.6 -1.0  0.6 -1.0 -0.2 -0.2
##  [31] -1.0  0.6 -1.4  0.2 -1.0 -0.2 -0.6  1.8 -0.6  0.2 -0.6 -1.4 -1.0  1.8 -0.2
##  [46]  1.0  1.4 -0.6 -0.2  1.0 -1.4  1.0  0.2 -1.0  1.0  0.6  0.6  0.2  1.0 -1.0
##  [61] -1.0  1.0 -0.6 -1.0  0.6  1.0 -0.2  0.6  0.2 -0.2 -0.2  0.2 -0.6 -0.6  0.6
##  [76] -0.2  1.0  1.4  1.8 -0.2  0.6 -1.0 -0.6  2.2  1.0  1.0  1.0  0.2 -1.4 -0.2
##  [91]  1.4  0.6 -0.2  0.6  0.6  2.2 -0.6 -1.0  0.6 -0.2
##
## $plot
```

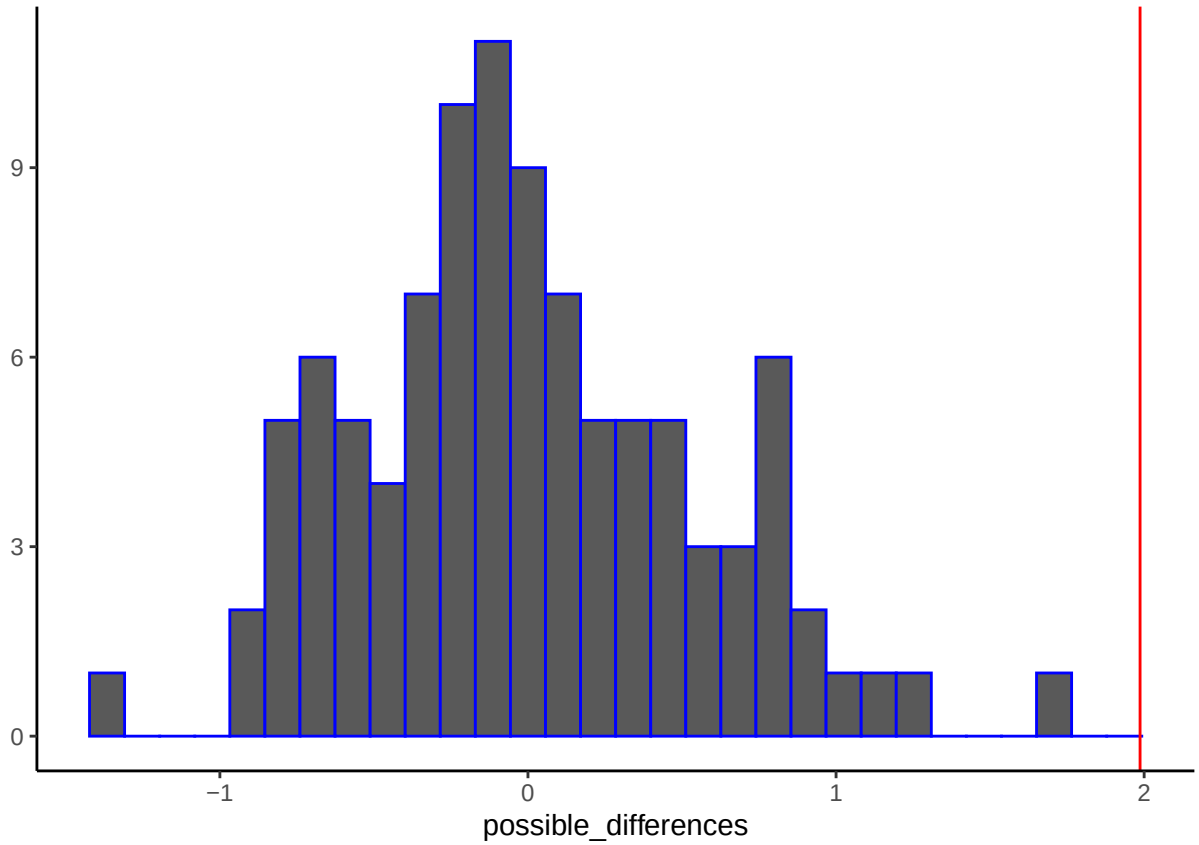



```
##
```

```
## $p_value
## [1] 0.21
```

```
the_data <- read.csv("open_data/SchroederEpley2015data.csv", header = TRUE)

group_0 <- the_data %>%
  filter(CONDITION == 0) %>%
  select(Intellect_Rating)

  group_0 <- group_0$Intellect_Rating

group_1 <- the_data %>%
  filter(CONDITION == 1) %>%
  select(Intellect_Rating)

group_1 <- group_1$Intellect_Rating

randomization_test(A = group_1,
                   B = group_0,
                   resample=100)
```

```
## $A_mean
## [1] 5.634921
##
## $B_mean
## [1] 3.648148
##
## $difference
## [1] 1.986772
##
## $randomization
##   [1] -0.592592593 -0.076719577  0.783068783 -0.420634921 -0.798941799
##   [6]  0.404761905  1.023809524 -0.867724868 -0.145502646  0.198412698
##  [11] -0.661375661  0.232804233 -0.214285714 -0.214285714  0.129629630
##  [16]  0.611111111 -0.592592593 -0.661375661  0.301587302  0.542328042
##  [21]  0.783068783 -1.314814815 -0.111111111 -0.420634921  0.129629630
##  [26] -0.076719577 -0.007936508  0.026455026  1.746031746 -0.626984127
##  [31] -0.111111111 -0.386243386 -0.386243386 -0.283068783  0.783068783
##  [36] -0.489417989  0.164021164  0.679894180  0.851851852  0.026455026
##  [41] -0.902116402  0.920634921 -0.592592593 -0.764550265 -0.248677249
##  [46] -0.076719577 -0.351851852 -0.111111111  0.335978836  0.095238095
##  [51]  0.232804233 -0.455026455  0.439153439 -0.007936508  0.920634921
##  [56]  0.404761905  0.129629630  0.714285714  0.542328042 -0.179894180
##  [61] -0.695767196 -0.145502646  0.507936508  0.060846561  0.335978836
##  [66]  0.748677249 -0.076719577 -0.042328042  0.404761905 -0.730158730
##  [71] -0.764550265  0.783068783 -0.523809524 -0.386243386 -0.179894180
##  [76] -0.214285714  1.092592593 -0.764550265  0.370370370 -0.386243386
##  [81] -0.042328042  0.026455026  0.714285714  0.198412698 -0.214285714
##  [86]  1.230158730 -0.798941799 -0.007936508  0.060846561  0.370370370
##  [91] -0.317460317 -0.042328042 -0.386243386 -0.076719577 -0.076719577
##  [96] -0.695767196 -0.248677249 -0.523809524 -0.248677249  0.267195767
##
## $plot
```




```
##
## $p_value
## [1] 0
```

The mean difference was 1.98 with a probability of 0. I would suggest that having a mean difference is actually not due to chance alone because i have a p value of 0 (but we should never say 0 so in this case it just means really really small or less than .01)