BIOST 544 HW2

Ivy Zhang

10/27/2021

Question 1

I wrote a function to simulate what a signle clinical trial will run like and get the difference of the proprtion of oucome = 1 between the treatment group and the contro group. Then I wrote another function to run this simulation of one clinical trial nsim times with nsim defined from the function definition. We will calculate what proportion of the simulation differences is more extreme than our observation in our real data, returning the tail probability. The function also generate a graph shows about the distribution of the simulated difference between the treatment group.

```
library(ggplot2)
simulate_single_trial = function(data){
  perm = sample(1:nrow(data),replace =FALSE)
  perm_data = data
  perm_data$tx = data$tx[perm]
  perm_mean_diff = with(perm_data, mean(outcome[tx==1])-mean(outcome[tx==0]))
  return(perm_mean_diff)
}
simulate_multi_trial = function(nsim, data){
  permuted_stats= replicate(nsim, simulate_single_trial(data))
  data_stats = with(data, mean(outcome[tx==1])-mean(outcome[tx==0]) )
  permu_mean_p = min(mean(data_stats<=permuted_stats),
                     mean(data_stats>=permuted_stats))
  permuted_stats = as.data.frame(permuted_stats)
  colnames(permuted_stats) = "stats"
  print(ggplot(permuted_stats, aes(x=stats, y=..density..)) +
    geom_density() +
    geom_vline(xintercept=data_stats, colour = "red"))
  return(permu_mean_p)
}
```

Question 2(a)

I first wrote a function to simulate an adaptive randomization scenarion with a data. I shuffled the order of the original data to randomize the potential probability of assigned treament to each observation. For each observation, we will calculate a new probability of getting assigned to treatment absed on the previous observation outcomes. For each simulated trials, I record the difference of proportion of oucome = 1 in the two treatment groups. Then I generated another function to run this single-trial simulation nsim times(nsim defined in the function parameter), calculating the proportion of the results of the simulated differences are more extreme than what we have in the real data as the tail probability. The function also generates a graph to show the distribution of the simulated differences in proportion.

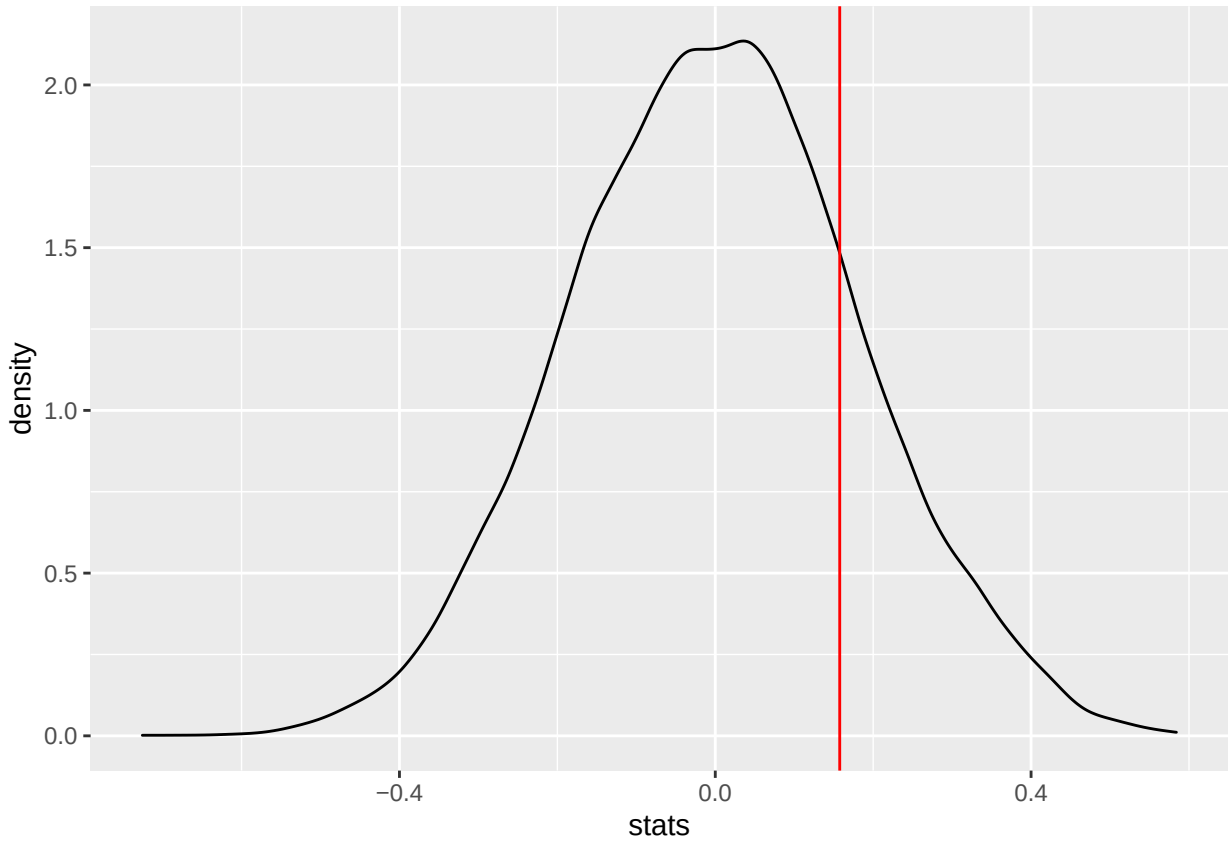
```
random_single_trial = function(data){
  n_succ = 0
  o_fail = 0
```

```
  p_new = 0.5
  ran_data = matrix(NA, nrow = nrow(data), ncol = 2)
  for(i in 1:nrow(data)){
    tx = rbinom(1,1,p_new)
    outcome = data$outcome[data$order == i]
    if(tx == 1&outcome ==1) {
      n_succ = n_succ+1
    }
    if(tx==0&outcome==0){
      o_fail = o_fail+1
    }
    ran_data[i,1] = tx
    ran_data[i,2] = outcome
    p_new = (1+3*(n_succ+o_fail))/(2+3*i)
  }
  colnames(ran_data) = c("tx","outcome")
  ran_data = as.data.frame(ran_data)
  ran_mean_diff = with(ran_data, mean(outcome[tx==1]-mean(outcome[tx==0])))
  return(ran_mean_diff)
}
random_multi_trial = function(nsim, data){
  permuted_stats= replicate(nsim, random_single_trial(data))
  data_stats = with(data, mean(outcome[tx==1])-mean(outcome[tx==0]) )
  permu_mean_p = min(mean(data_stats<=permuted_stats),
                     mean(data_stats>=permuted_stats))
  permuted_stats = as.data.frame(permuted_stats)
  colnames(permuted_stats) = "stats"
  print(ggplot(permuted_stats, aes(x=stats, y=..density..)) +
    geom_density() +
    geom_vline(xintercept=data_stats, colour = "red"))
  return(permu_mean_p)
}
```

Question 2(b) & Question 3

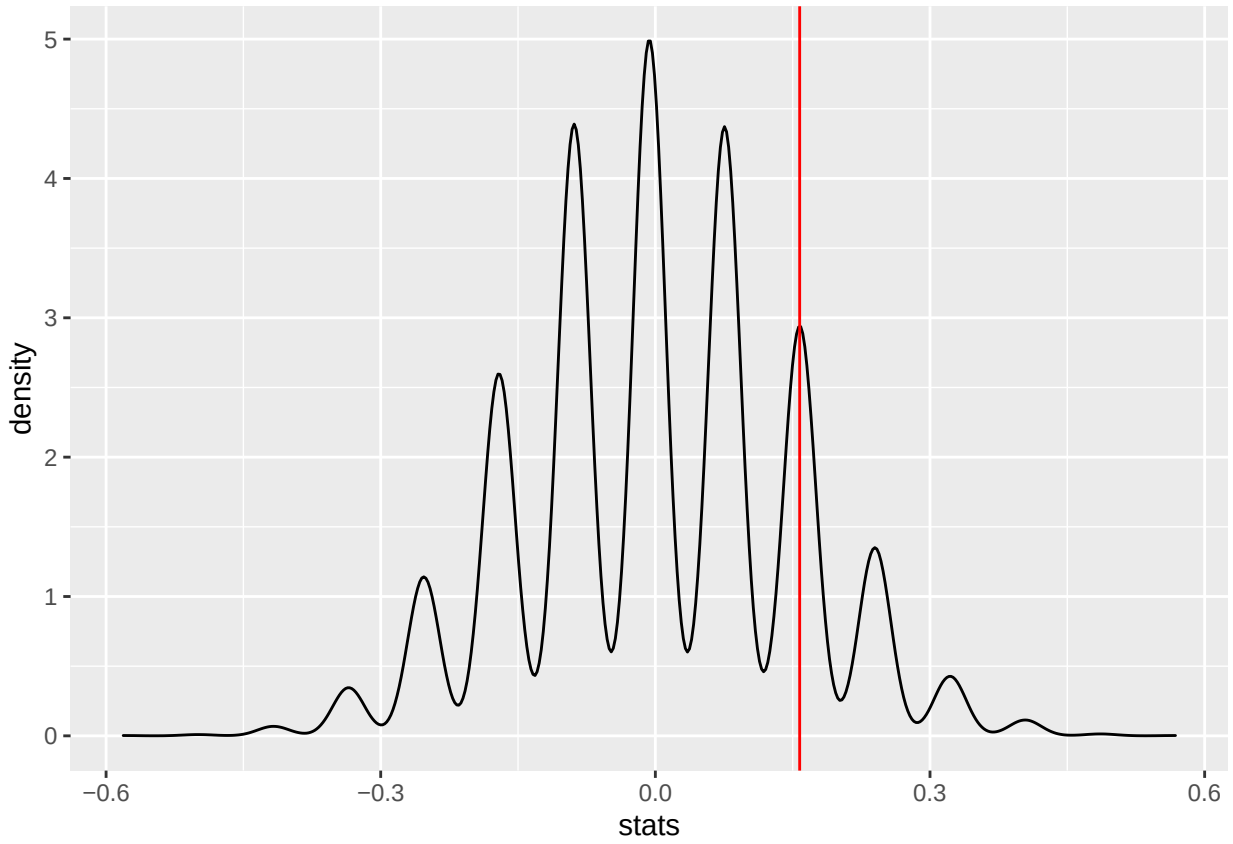
For the rest two sections, I called the function of running multiple simulations function I wrote in the previous sections with nsim = 10000 and data from the assignments.

```
dat <- read.csv("~/Desktop/R hw/HW2-adaptive-trial.txt")
nsim = 10000
set.seed(1)
random_multi_trial(nsim,dat)
```




```
## [1] 0.22
```

```
simulate_multi_trial(nsim, dat)
```




```
## [1] 0.2131
```

Based on the results, we can see the tail probability from the simulation of adaptive randomization is about 0.22. We cannot reject the hypothesis that the standard-of-care is at least as effective as the new treatment at the significance test of 0.05. However, it seems relatively unlikely for our data to have such difference when the standard-of-care is at least as effective as the new treatment.

The simulated distribution of the difference without knowing the trial is adaptive randomization trial seems to be similar to the distribution from the previous distribution, giving the tail probability around 0.21. Same as before, we cannot reject the hypothesis that the standard-of-care is at least as effective as the new treatment at the significance test of 0.05. These two distribution have similar range and variance, and both have peak around 0. Although the graph of the later distribution seems to looked not exactly like bell-shape, I think it is due to the sample size and having the two treatment groups sizes fixed. Overall, I think they have similar distribution.