ETC 2420/5242 Lab 3 2017

Di Cook

SOLUTION

Purpose

This lab is to examine testing hypotheses using permutation.

Background

Read Sections 2.3, and the exercises in 2.9.3, of the online textbook "IntroStat with Randomization and Simulation".

Problem description

Is yawning contagious? An experiment conducted by the MythBusters, a science entertainment TV program on the Discovery Channel, tested if a person can be subconsciously influenced into yawning if another person near them yawns. 50 people were randomly assigned to two groups: 34 to a group where a person near them yawned (treatment) and 16 to a group where there wasn't a person yawning near them (control). The following table shows the results of this experiment.

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

Question 1 (5 pts)

a. How many subjects participated in the experiment? `50`
b. How were participants assigned to treatment and control groups? `Randomization`
c. What are the two variables that describe the experiment? `Group, yawn`
d. Compute the proportion of the treatment and control groups who yawned. Add this to the table. 0.25
e. Compute the difference in proportions between the two groups. `Control-Treatment is` -0.044

Question 2 (3 pts)

The null hypothesis for the experiment is

$$H_o : p_{control} = p_{treatment}$$

a. Write the null hypothesis as an English sentence. `Yawning is NOT contagious`
b. What would be the alternative hypothesis tested by MythBusters? $H_a : p_{control} < p_{treatment}$
c. Explain your reasoning. `The original question, or the common belief is that yawning is contagious, which would correspond to a higher proportion of people yawning in the group with the yawning near them.`

Question 3 (4 pts)

Write a function that permutes the `yawn` variable, and computes the difference between proportions of treatment and control groups.

```
prop_dif <- function(dat) {
  dtbl <- dat %>%
    mutate(yawn=sample(yawn)) %>%
    group_by(group, yawn) %>%
    tally() %>%
    ungroup() %>%
    spread(yawn, n, fill=0) %>%
    mutate(total = rowSums(.[-1])) %>%
    mutate(p = yes/total)
  return(pdif=dtbl$p[2]-dtbl$p[1])
}
```

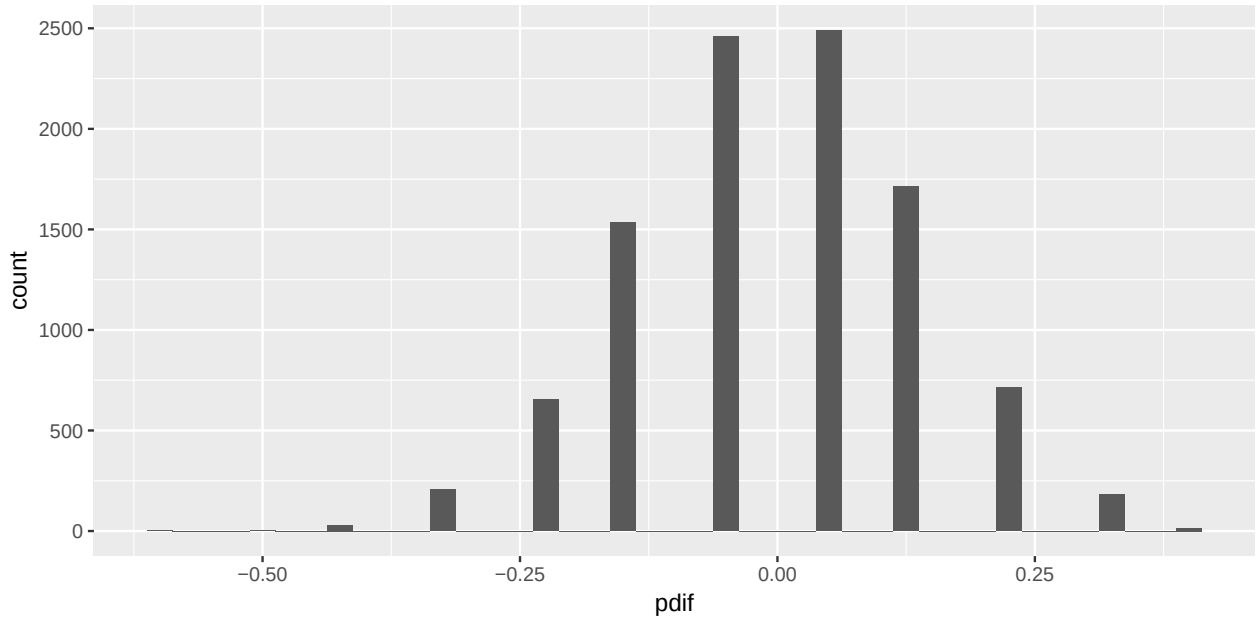
Question 4 (4 pts)

a. Run the function 10000 times, saving the result.

```
set.seed(444)
pdif <- NULL
for (i in 1:10000)
  pdif <- c(pdif, prop_dif(yawn_expt))
```

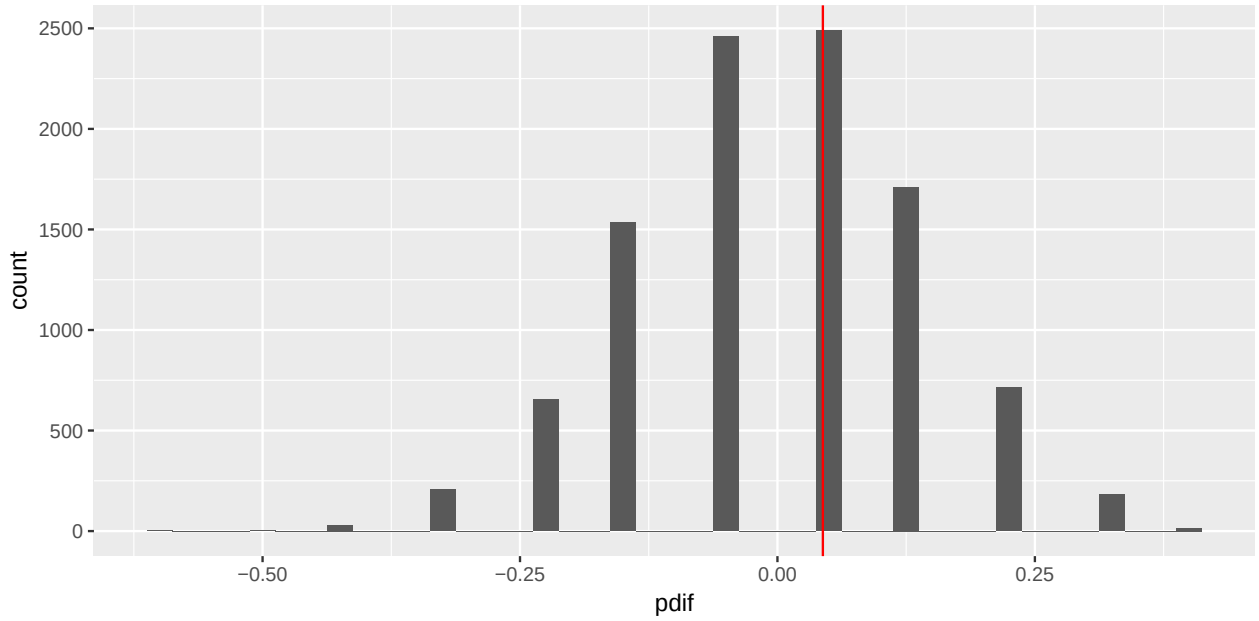
b. Make a histogram (or a dotplot) of the results.

```
library(ggplot2)
pdif <- data.frame(pdif)
ggplot(pdif, aes(x=pdif)) + geom_histogram(binwidth=0.025)
```



c. Draw a vertical line on the plot that represents the difference for the actual data.

```
ggplot(pdif, aes(x=pdif)) + geom_histogram(binwidth=0.025) +
  geom_vline(xintercept=0.0441176, colour="red")
```



d. Compute the proportion of times that the permuted data yields a difference larger than the difference of the actual data.

```
length(pdif[pdif>0.0441176])/10000
# [1] 0.5112
```

Question 5 (4 pts)

a. Compute the (permutation) p-value for testing the null hypothesis. 0.5112
b. Based on your p-value, what is your decision about the null hypothesis? `Fail to reject the null`
c. Write a sentence stating your conclusion. `There is no difference between the proportion of people yawning in the treatment and control groups.`
d. Finally, based on these experimental results how would you answer "Is yawning contagious?" `There is no evidence from this study to suggest that yawning is contagious.`