STA130 Final Project

How the launch of request expiry changed user behavior

Haiyue Yang, Yijing Chen, Xi Zheng(TUT0103-04)

Introduction

On November 21, 2018, Riipen, a Vancouver-based education technology company, launched a new feature called “request expiry”. This project is going to show how user behaviour have changed after the launch of request expiry from the following perspectives.

Objectives

Change in the behaviour of recipients

- Whether the **response rate** (in 14days) changed?
- How the **response time** has changed?

Change in the behaviour of creators

- How the **number of requests created per day** has changed?
- How the **number of accounts created per month** has changed?

Data Summary

- mutate()&ifelse(): add the type of users into request.xlsx
- mutate()&as.Date(): change variable types from chr to date
- mutate()&substr(): add the month of each request created at
- filter(): filter out data on 2018-4-12 & 2018-8-30 (extreme)

Response rate (in 14 Days)

State hypothesis

Null hypothesis: $H_0 : rate\_diff = 0$
The response rate did not change after the November 21 launch of request expiry.
Alternative hypothesis: $H_a : rate\_diff \neq 0$
The response rate has changed after the November 21 launch of request expiry.

Calculate test statistic:

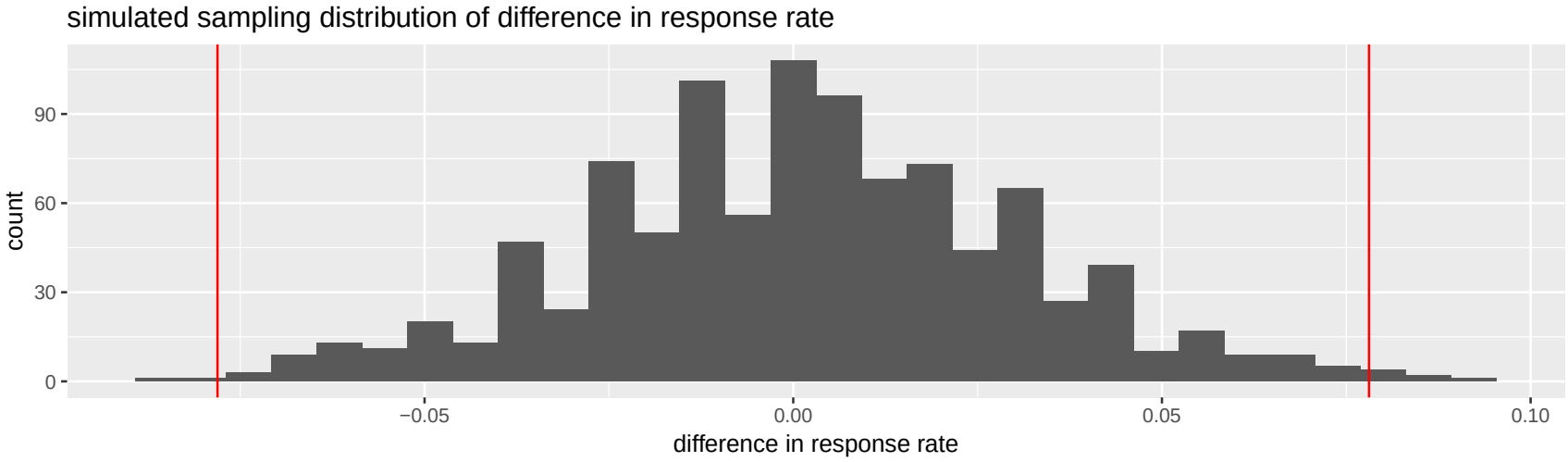
- Filter out the requests still pending
- Summarise the response rate before and after 2018-11-21
- Calculate the difference.
- Test statistic (rate_diff):

```
## [1] 0.07808326
```

Response rate (in 14 Days)

Simulate under the null hypothesis

- Shuffle the variable of the day requested created at
- Calculate the difference between the response rate before and after 2018-11-21
- Repeat 1000 times
- Turn the result into a data frame and plot the data



Response rate (in 14 Days)

Assess evidence against the null hypothesis

- Compare the test statistic to the simulated sampling distribution
- Calculate the p-value:

```
## [1] 0.009
```

Conclusion

p-value=0.009<0.01
We have strong evidence against the null hypothesis

Response rate (in 14 Days)

Different types of recipients

professor: too few observations to draw a conclusion
student:

- test statistic: stu_diff=0.1768805
- p-value=0
- 95% confidence interval

```
##        2.5%       97.5%
## 0.07873107 0.27436534
```

employer:

- test statistic: emp_diff=0.01536152
- p-value=0.706>0.1

Response time

hypothesis test & p-value

- hypothesis- hypothesis
  Null hypothesis:
  The average response time did not change after the November 21 launch of request expiry.

$$H_0 : time\_diff = 0$$

  Alternative hypothesis:
  The average response time has changed after the November 21 launch of request expiry.

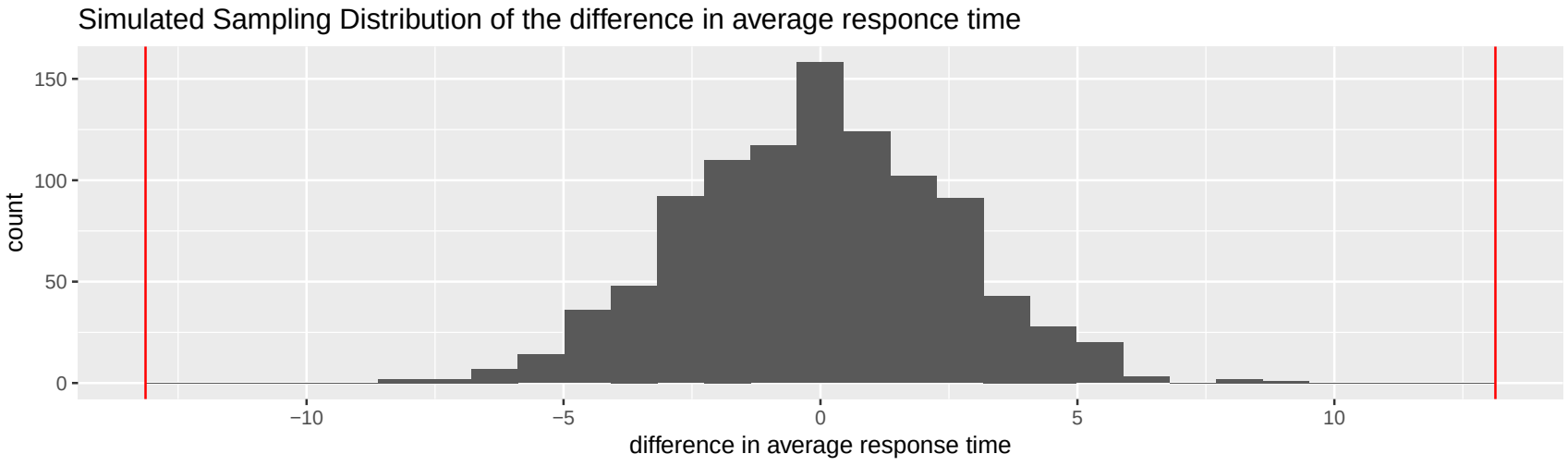
$$H_a : time\_diff \neq 0$$

- test statistic: (time_diff)

```
## [1] -13.13395
```

Response time

hypothesis test & p-value



- p-value

```
## [1] 0
```

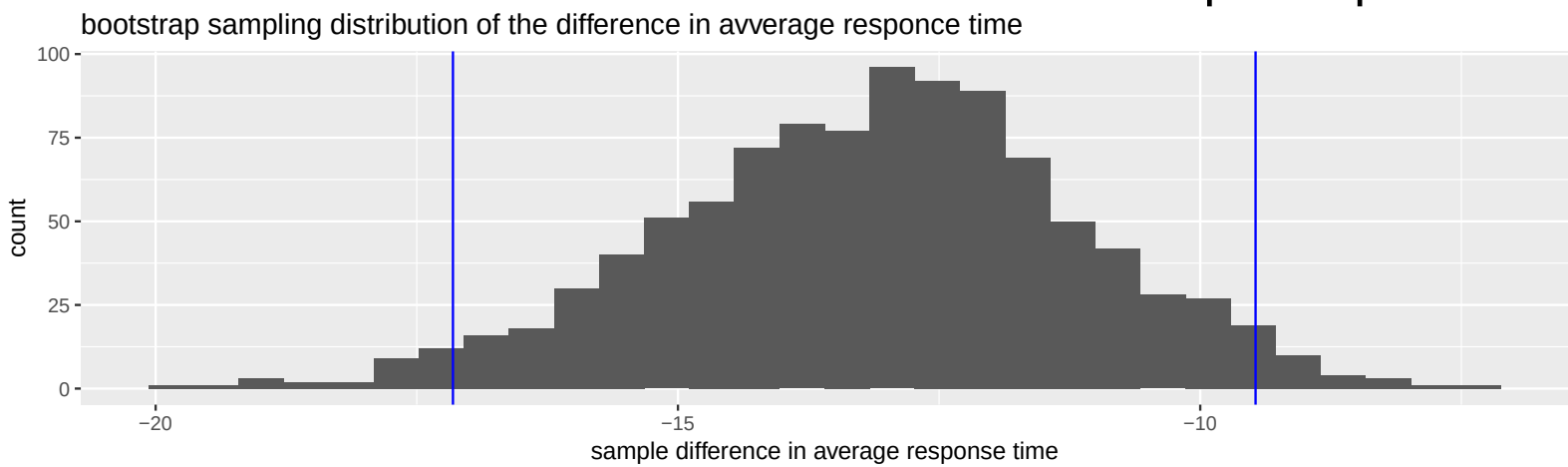
- conclusion
  p-value=0
  We have strong evidence against the null hypothesis.

Response time

bootstrap sampling & confidence interval

bootstrap sample

- Draw 1000 bootstrap samples of the same size as the original sample with replacement
- For each bootstrap sample, calculate the statistic
- Summarise the value of statistic for all bootstrap samples



Response time

bootstrap sampling & confidence interval

95% confidence interval

- Calculate the middle 95% interval of values of the bootstrap statistic

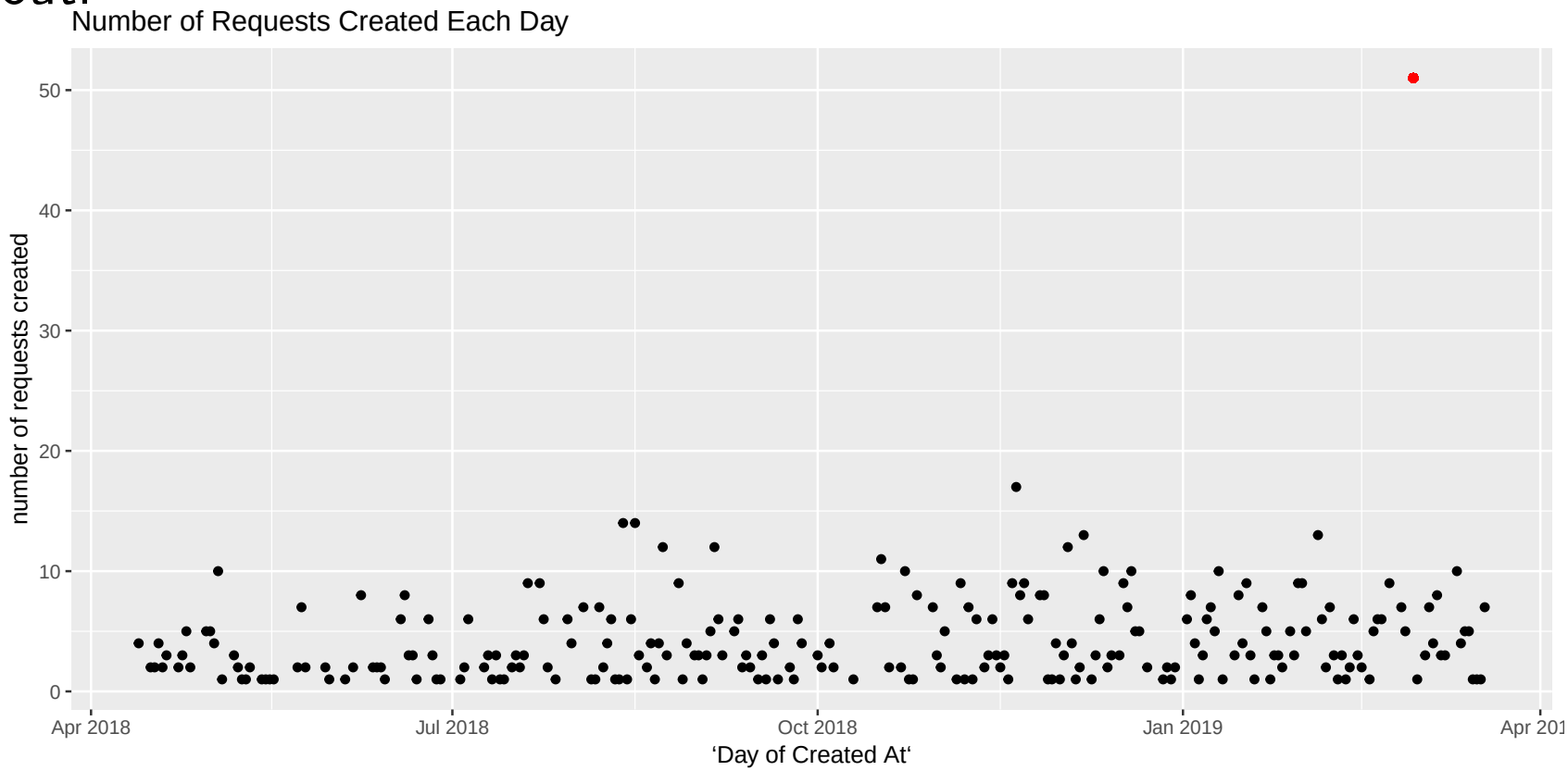
```
##       2.5%      97.5%
## -17.154014  -9.470533
```

We are 95% confident that the response time has decreased by between 9.470533 and -17.154014 days after 2018-11-21.

Number of requests per day

clear out outliers

We plotted a scartter plot for number of requests created each day. From the plot, we discovered an outlier at 2019-02-28 and cleared it out.



Number of requests per day

hypothesis test & p-value

- hypothesis
  Null hypothesis:
  The number of requests created per day did not change after the November 21 launch of request expiry.

$$H_0 : num\_diff = 0$$

  Alternative hypothesis:
  The number of requests created per day has changed after the November 21 launch of request expiry.

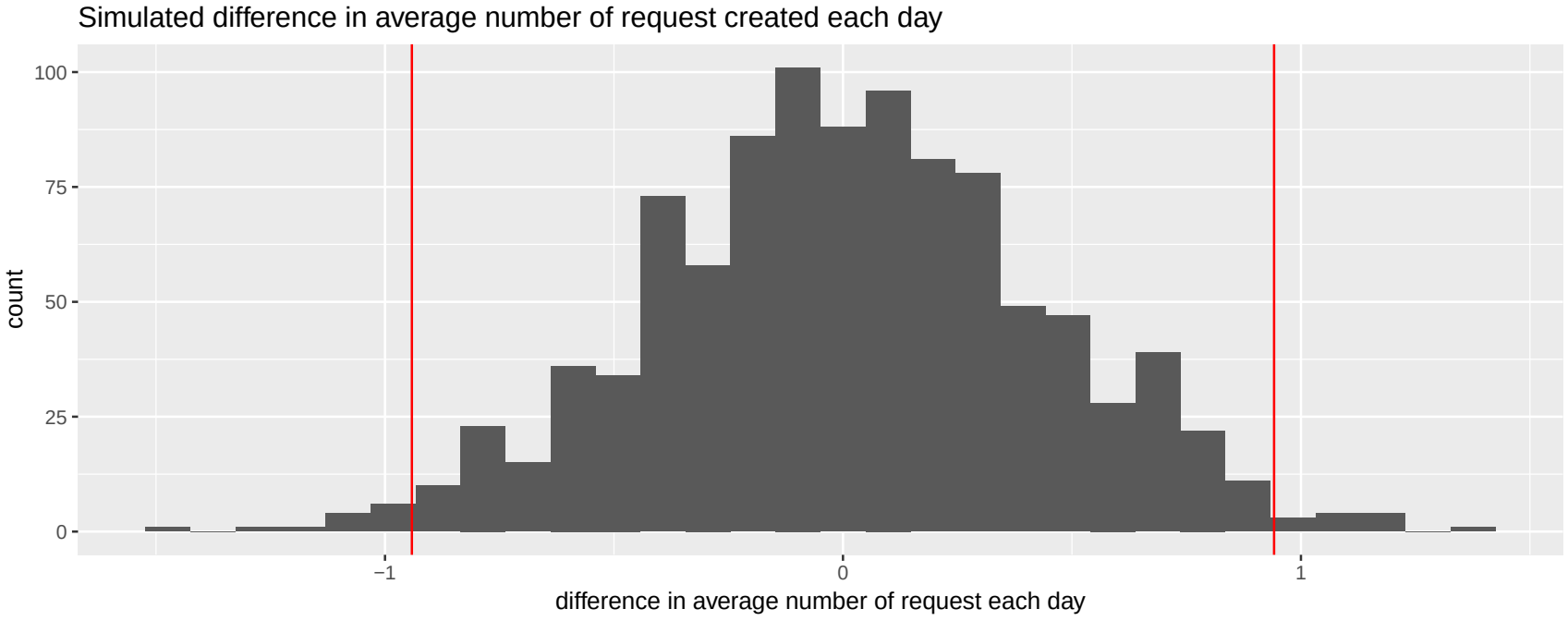
$$H_a : num\_diff \neq 0$$

- test statistic: (num_diff)

```
## [1] 0.9413082
```

Number of requests per day

hypothesis test & p-value



- p-value

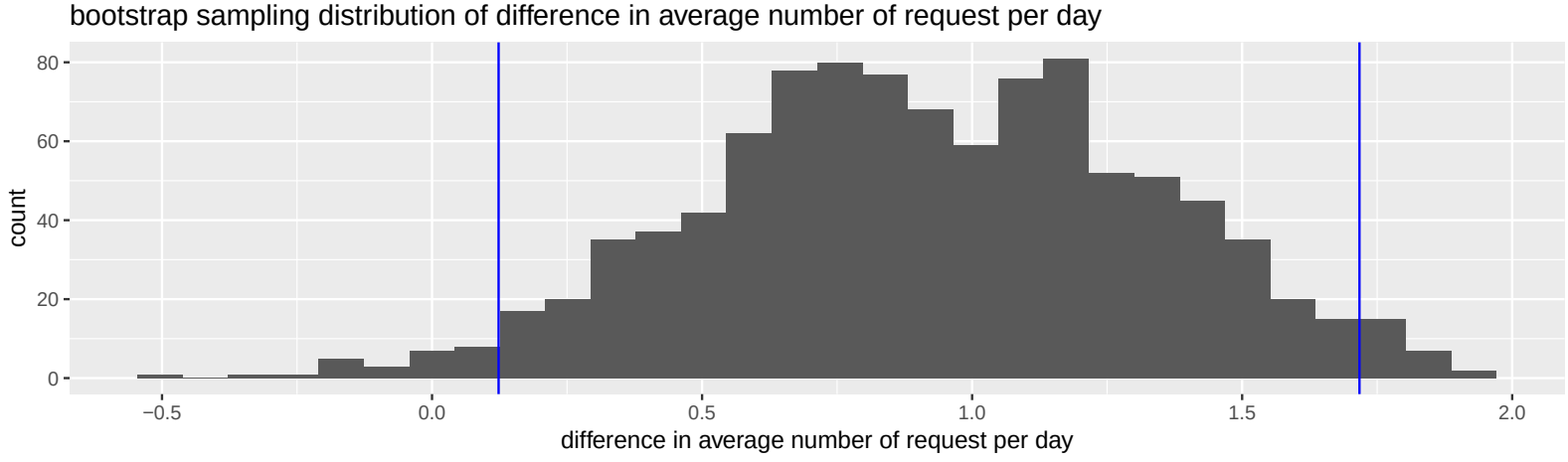
```
## [1] 0.024
```

- conclusion
  p-value=0.024<0.05
  We have evidence against the null hypothesis.

Number of requests per day

bootstrap sampling & confidence interval

- bootstrap sampling distribution:



- 95% confidence interval

```
##       2.5%      97.5%
## 0.1232033 1.7172502
```

Number of new accounts created

clean up data

- summarise the number of accounts craeted each month
- filter out accounts created on ‘2017-02-11’ because it is the start date.
- filter out accounts created in ‘2019-03’ because the month is not complete.

Number of new accounts created

hypothesis test & p-value

- hypothesis
  Null hypothesis:
  The number of accounts created each month did not change after the November 21 launch of request expiry.

$$H_0 : account\_diff = 0$$

  Alternative hypothesis:
  The number of accounts created each month has changed after the November 21 launch of request expiry.

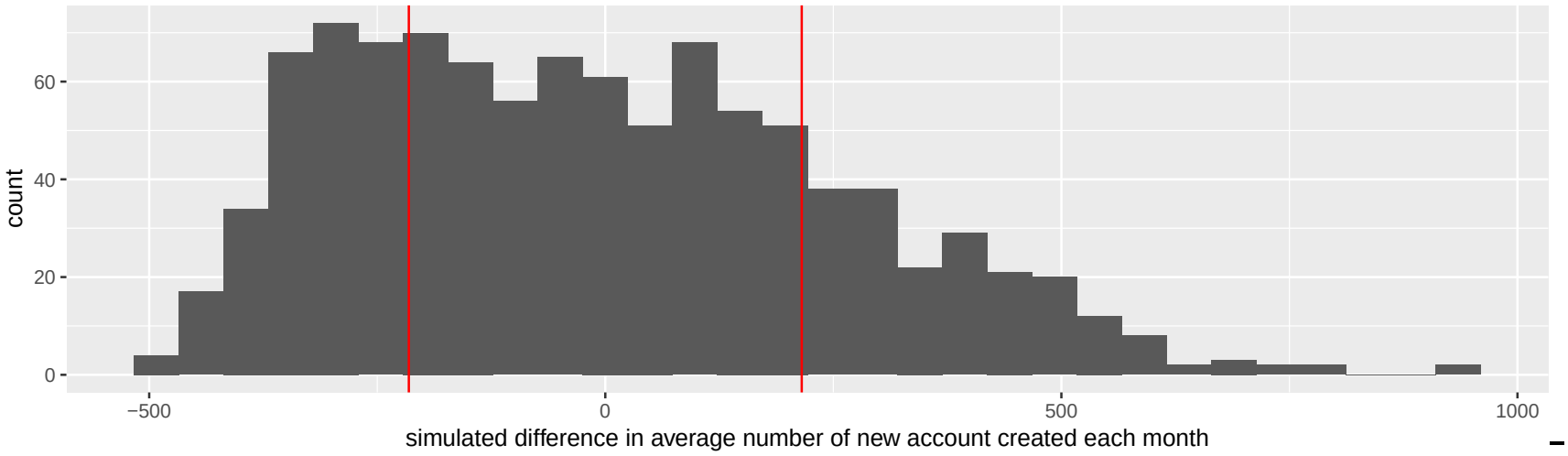
$$H_a : account\_diff \neq 0$$

- test statistic (account_diff):

```
## [1] 215.3636
```

Number of new accounts created

hypothesis test & p-value



p-value

```
## [1] 0.476
```

- conclusion p-value=0.476>0.1 We do not have evidence against the null hypothesis.

Conclusion

Change in Behaviour of Recipients

response rate

- The response rate has changed after the launch of request expiry.
- More specifically:
  the response rate of employers did not change; the response rate of students has changed, and we are 95% confident that it is increased by between 0.07873107 and 0.27436534.

response time

- The average response time has changed after the launch of request expiry.
- We are 95% confident that the average responce time has decreased by between 9.470533 and 17.154014 days.

Conclusion

Change in Behaviour of Creators

number of requests each day

- The number of requests created each day has changed after the launch of request expiry.
- We are 95% confident that the average number of requests per day has increased by between 0.1232033 and 1.7172502

number of new accounts each month

- The number of new accounts created each month did not change after the launch of request expiry.

Conclusion

possible explanations

- The launch of request expiry has increased the response rate within 14 days and decreased the responses time, and thus improved user experience.
- Thus,users tend to use Riipen more.
- Since request expiry has launch only 4 months, it has not draw more new users, but is probably going to draw more users.

limitation

We did not analyze how many requests are recent after expiry, so probably the number of requests created increase simply because some requests are recent after expiry.