

STA130 Final Project

Riipen Data Analysis

Muxian Chen/Yuze Kang/Chenyang Pan/Yijie Zhao,
Section:L0102, Group Number:102-1

Introduction

Riipen is a Vancouver-based education technology company, which connects schools to employers to create project-based learning opportunities and also reduce underemployment through the experiential learning. On November 2018, a new feature “request expiry” has launched due to some slow response. We will then further analyze the effect of this feature.

Objectives

To fully address the problem of Riipen and estimate the most appropriate statistical methods from STA130, analyzing the data of Riipen in order to test whether adding a 'requests expiry' would affect users' behaviour.

Data Cleaning Process

- ▶ Files used: requests, employeraccounts, educatoraccounts, studentaccounts.
- ▶ Outliers: requests on Aug 30, 2018.

Methods

- ▶ Two Proportion Hypothesis Test
- ▶ Barplot
- ▶ Boxplot
- ▶ Bootstrap Sampling Distribution

1.1 Number of Requests per User

► Employers

```
## Observations: 448
## Variables: 2
## $ actor_id <dbl> 718, 2887, 4802, 4923, 5447, 5500, 5562
## $ n          <int> 1, 1, 1, 1, 1, 1, 1, 2, 1, 6, 1, 1, 1,
```

► Educators

```
## Observations: 93
## Variables: 2
## $ actor_id <dbl> 4109, 8325, 8552, 8554, 8578, 8605, 911
## $ n        <int> 9, 17, 2, 2, 1, 3, 2, 2, 6, 1, 2, 2, 1,
```

1.2 Proportion of Requests from Educators and Employers

```
## # A tibble: 4 x 3
## # Groups:   created_after_1121 [2]
##   created_after_1121 requestable_model      n
##   <chr>                <chr>          <int>
## 1 After11-21          course          210
## 2 After11-21          project         250
## 3 Before11-21         course          160
## 4 Before11-21         project         567
```

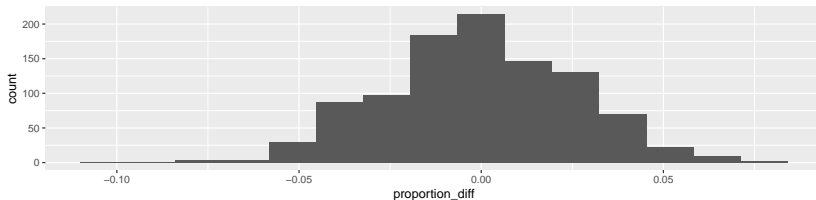
1.2 Proportion of Requests from Educators and Employers

- ▶ **Null Hypothesis(H_0):** Proportion of requests from educator accounts does not change after setting expire date. ($P_{after} - P_{before} = 0$)
- ▶ **Alternative Hypothesis(H_A):** Proportion of requests from educator accounts changes after setting expire date. ($P_{after} - P_{before} \neq 0$)
- ▶ **test statistics**

```
## [1] 0.2364392
```

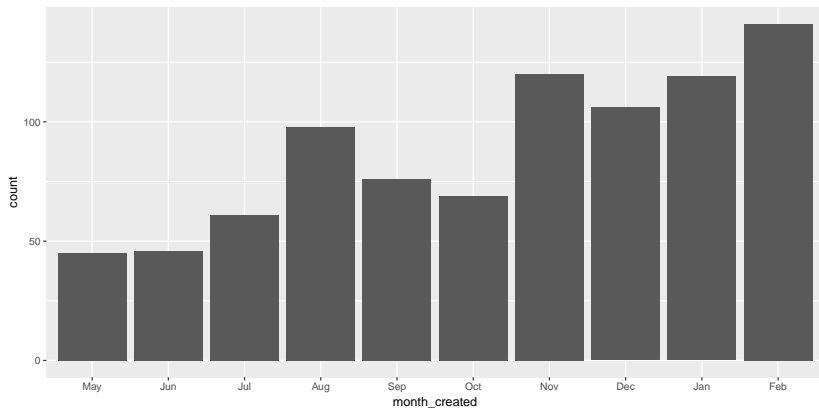

1.2 Proportion of Requests from Educators and Employers

Hypothesis Test

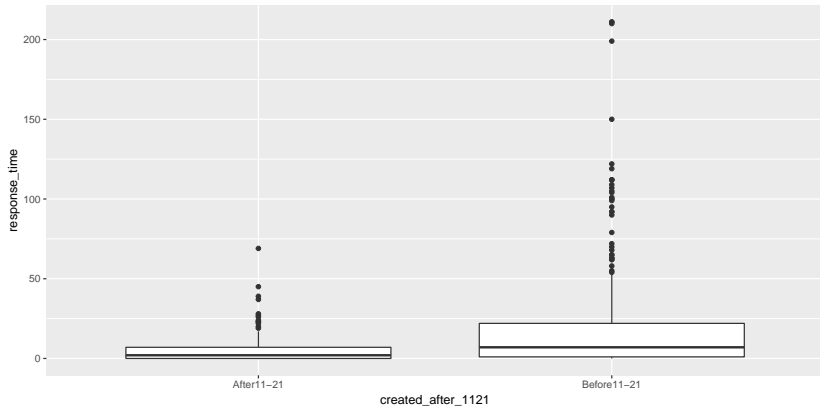


```
## # A tibble: 1 x 1
##   p_value
##   <dbl>
## 1      0
```

2. Volume of Requests Over Time



3. Response Time (Rate)



4. Acceptance Rate Before and After 2018-11-21

Number of requests of different state

► Before 2018-11-21

```
## # A tibble: 4 x 2
##   state      n
##   <chr>    <int>
## 1 accepted    279
## 2 cancelled   221
## 3 expired     53
## 4 rejected   174
```

4. Acceptance Rate Before and After 2018-11-21

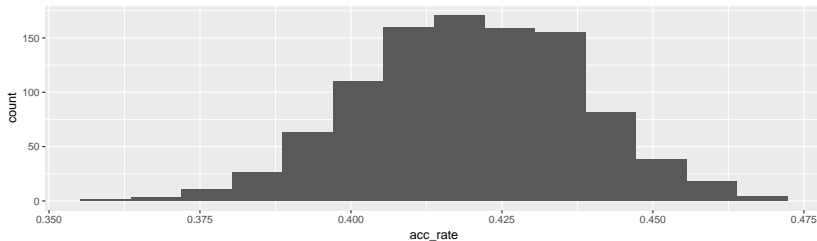
Number of requests of different state

► After 2018-11-21

```
## # A tibble: 5 x 2
##   state      n
##   <chr>    <int>
## 1 accepted    219
## 2 cancelled    33
## 3 expired    116
## 4 pending     26
## 5 rejected     66
```

4.1 Acceptance Rate Before 2018-11-21

Bootstrap Sampling Distribution

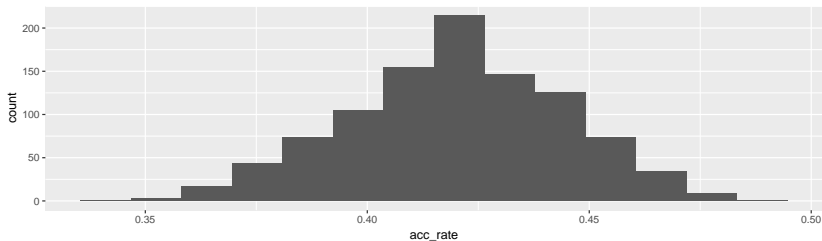


2.5% 97.5%

0.3851100 0.4552957

4.2 Acceptance Rate After 2018-11-21

Bootstrap Sampling Distribution

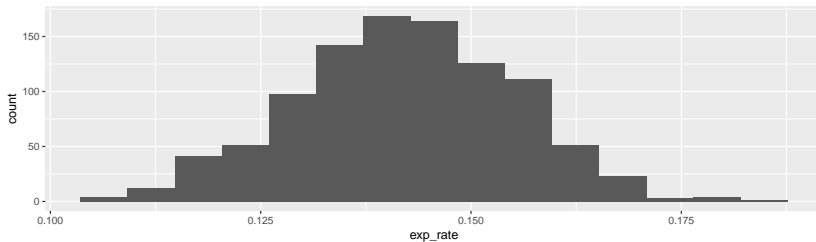


2.5% 97.5%

0.3717391 0.4673913

5.1 Expire Rate Before 11-21

Bootstrap Sampling Distribution

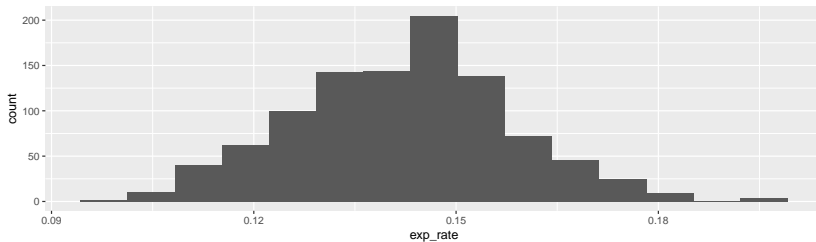


2.5% 97.5%

0.1169188 0.1664718

5.2 Expire Rate After 11-21

Bootstrap Sampling Distribution



2.5% 97.5%

0.1130435 0.1739130

Conclusion

After setting requests expiry

- ▶ send more requests
- ▶ respond more quickly
- ▶ proportion of requests from education increases
- ▶ true acceptance rate does not change
- ▶ true expire rate does not change

Limitation