A/B Test with Python

Yasemin Turker

2022

INTRODUCTION

Problem

An advertising company has developed a new ad to have users engage with their questionnaire.

The company has shown the new ad to some users and a dummy ad to others and wants their data analyst team to interpret the results.

Does the new ad generate more responses to their questionnaire?

Is it statistically significant?

Is the company justified in using the new ad?

"

A/B testing is a type of split testing and is commonly used to drive improvements to specific variables or elements by measuring user or audience engagement.[1]

DATA

Dataset is found on kaggle from an advertising company:

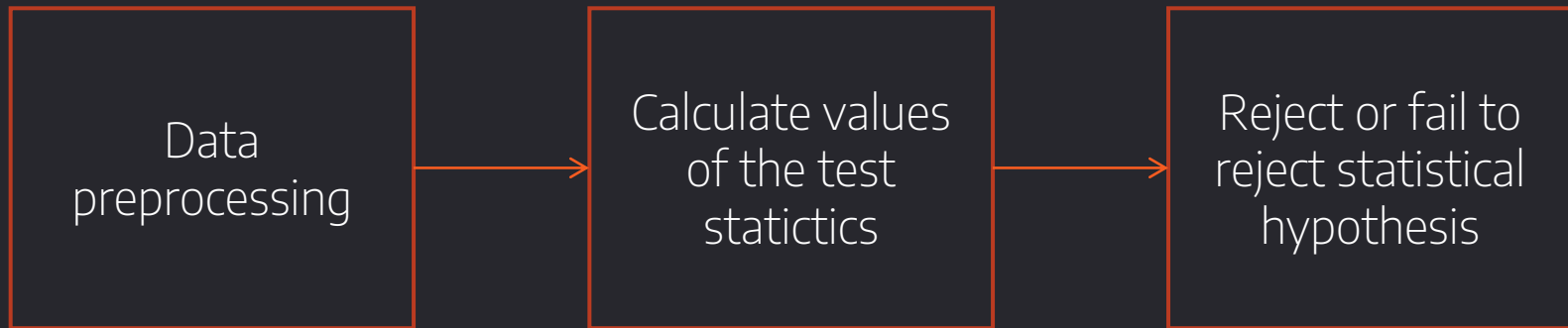https://www.kaggle.com/osuolaleemmanuel/ad-ab-testing

$$H_0: \mu_{con} = \mu_{exp}$$

Null Hypothesis : There is no significant difference between the ad success rate of both groups

$$H_1: \mu_{con} \neq \mu_{exp}$$

Alternative Hypothesis : There is significant difference between the ad success rate of both groups

A/B TESTING PROCESS

Data preprocessing → Calculate values of the test statictics → Reject or fail to reject statistical hypothesis

DATA PREPROCESSING

- Check for null values

- Remove non-answer observations

DATA PREPROCESSING

Checked for null values


```python
# Checking for nulls
Total_nulls = ABdata.isnull().sum()
print('Null values:')
print(Total_nulls)
```

```
Null values:
auction_id      0
experiment      0
date            0
hour            0
device_make     0
platform_os     0
browser         0
yes             0
no              0
dtype: int64
```

DATA PREPROCESSING

Removed non-answer observations

DATA PREPROCESSING

Removed non-answer observations

```
# dropped non-answer observations
ABdata_A00= ABdata.drop(ABdata_00.index)
print('# Observations: {}'.format(ABdata_A00.shape[0]))
ABdata_A00.head()
```

Observations: 1243

	auction_id	experiment	date	hour	device_make	platform_os	browser	yes	no
2	0016d14a-ae18-4a02-a204-6ba53b52f2ed	exposed	2020-07-05	2	E5823	6	Chrome Mobile WebView	0	1
16	008aafdf-deef-4482-8fec-d98e3da054da	exposed	2020-07-04	16	Generic Smartphone	6	Chrome Mobile	1	0
20	00a1384a-5118-4d1b-925b-6cdada50318d	exposed	2020-07-06	8	Generic Smartphone	6	Chrome Mobile	0	1
23	00b6fadb-10bd-49e3-a778-290da82f7a8d	control	2020-07-08	4	Samsung SM-A202F	6	Facebook	1	0
27	00ebf4a8-060f-4b99-93ac-c62724399483	control	2020-07-03	15	Generic Smartphone	6	Chrome Mobile	0	1

EXPLORATORY DATA ANALYSIS

- Process of performing some initial investigations on the dataset

EXPLORATORY DATA ANALYSIS

```
NumberofControl=ABdata_A00['experiment'].value_counts().control
NumberofExposed=ABdata_A00['experiment'].value_counts().exposed
NumberofControl,NumberofExposed
```

```
(586, 657)
```

```
Control_yes= ExperimentsYes.iloc[0]
Exposed_yes= ExperimentsYes.iloc[1]
Control_yes,Exposed_yes
```
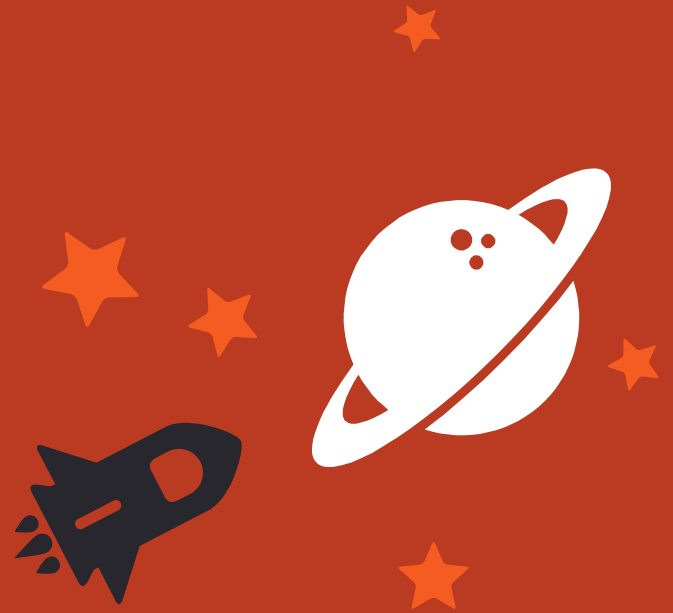
```
(264, 308)
```

```
RateOfControl= Control_yes/NumberofControl
RateOfExposed= Exposed_yes/NumberofExposed
RateOfControl, RateOfExposed
```

```
(0.45051194539249145, 0.4687975646879756)
```

2 SAMPLE Z TEST

- Investigate the significant differences between the control and experimental groups' metrics

2 SAMPLE Z-TEST

Test Stat.	Z Critical	P Value	Confidence Interval
-0.6457	1.9599	1.4815	[-0.07380 0.03722]

Conclusion
Test stat:0.6457 < Z Crit:1.96 can not be rejected the null hypothesis

CONCLUSIONS

- The null hypothesis can not be rejected and conclude that there is no statistically significant difference between the AdA and AdB.

- The company is not justified in using the new ad because the findings show that the new ad will not provide any additional benefit.

THANKS!

ANY QUESTIONS?

- yaseminaturker@gmail.com

REFERENCES

1- https://www.seldon.io/a-b-testing-for-machine-learning